9-2017

Simulation-based Estimation and Inference of Production Frontiers

Thomas Tao YANG

Yichong ZHANG
Singapore Management University, yczhang@smu.edu.sg

Simulation-based Estimation and Inference of Production Frontiers*

Thomas Tao Yang† Yichong Zhang‡

September 27, 2017

Abstract

This article proposes two novel estimation and inference approaches for production frontiers based on extreme quantiles of feasible outputs. The first approach linearly combines two extreme quantiles to reduce the estimation bias, and uses a subsampling method to construct point estimates and confidence intervals. The second approach can accommodate any finite number of extreme quantile estimates by way of the Approximate Bayesian Computation method. The point estimators and confidence intervals are then obtained through the Markov Chain Monte Carlo algorithm. The estimations and inferences of both approaches are justified asymptotically. Their finite sample performances are illustrated through simulations and an empirical application.

Keywords: Fixed-k Asymptotics, Extreme Value Theory

*First draft: September, 2017. We are grateful to Valentin Zelenyuk and the participants of 2017 HU-HUE-SMU Tripartite Conference on Econometrics for valuable comments.

†Australian National University. E-mail address: tao.yang@anu.edu.au.

‡Singapore Management University. E-mail address: yczhang@smu.edu.sg.

1 Introduction

The estimation of the production frontier (or data envelope) arises naturally in and applies to many fields such as manufacturing, health care, transportation, education, banking, public services, and portfolio management. A survey by Gattoufi, Oral, and Reisman (2004) listed over 1,400 references to the topic of data envelopment. However, the estimation and inference of production frontiers are complicated by the fact that the parameter of interest is on the boundary, and thus is non-regular.

In this article, we propose two novel estimation and inference approaches for the frontier. Our approaches are robust to certain amount of outliers as they are based on extreme quantiles, rather than the sample maximum, of feasible outputs. In addition, our approaches correct the downward bias between the extreme quantile and the production frontier by modern simulation-based methods, which avoid analytically estimating the bias. Consequently, our estimators and the followed inferences are not contaminated by errors from the bias estimation.

Our first approach uses a linear combination of two extreme quantiles to estimate the frontier, and a subsampling method to construct point estimators and confidence intervals. We refer to it as the subsampling approach. The second approach is able to utilize any finite number of extreme quantiles. It treats the extreme quantile estimates as new observations and approximates their likelihood using their joint asymptotic distribution. It then puts a prior on the production frontier, draws the posterior distribution by Markov Chain Monte Carlo (MCMC) method, and constructs point estimators and confidence intervals based on the posterior distribution. Because the likelihood is approximated, we refer to this approach as the Approximate Bayesian Computation (ABC) approach. We use a feasible normalizing factor in both approaches, without imposing additional restrictions. The inferences of both approaches are justified asymptotically.

Our article contributes to three branches of literature: the estimation and inference of production frontiers, the inference of extreme quantiles, and the ABC inference. We relate them one by one as follows.

A pioneering work by Deprins, Simar, and Tulkens (1984) first introduced the free-disposal hull (FDH) estimator of production frontiers. Its asymptotic properties have been studied by Park, Simar, and Weiner (2000) and Daouia, Florens, and Simar (2010). Given convexity of the production frontier, another important work by Kneip, Park, and Simar (1998) considered the data envelopment analysis (DEA) estimators. The asymptotic properties of DEA estimators have been investigated by Kneip et al. (1998), Gijbels, Mammen, Park, and Simar (1999), Jeong (2004), Jeong and Park (2006), Kneip, Simar, and Wilson (2008), and Park, Jeong, and Simar (2010). However, neither the FDH or DEA estimators are robust to any outliers. In addition, the inference of the FDH estimator requires estimating the convergence rate, while a valid inference for the DEA estimator is still lacking, to the best of our knowledge. Recognizing those drawbacks, Cazals, Florens, and Simar (2002) and Aragon, Daouia, and Thomas-Agnan (2005) suggested estimating an expected frontier, but it does not envelope the data. Daouia et al. (2010) and Daouia, Girard, and Guillou (2014) proposed using intermediate quantiles, and then extrapolating to the boundary. Various bias correction methods are proposed which require certain higher-order expansion of the tail distribution. Our approaches do not require the frontier to be convex, and thus are in spirit closer to the FDH estimator. We use extreme, rather than intermediate, quantiles to construct estimators and correct bias using simulation-based methods. Consequently, our approaches are robust to a few outliers and do not rely on any higher-order expansion assumption of the tail distribution.

The literature on the inference of extreme quantiles includes Bertail, Haefke, Politis, and White (2004), Chernozhukov and Fernández-Val (2011), and Zhang (2016), in the contexts of percentiles, linear quantile regressions, and quantile treatment effects, respectively. Our article complements the literature by studying the production frontier with a new feature of the estimand on the boundary. Our first approach conducts inference by subsampling, which is also the case in Simar and Wilson (2011) but for the FDH estimator. Unlike theirs, our subsampling inference approach is for extreme quantiles but not sample maximum, and does not require knowledge of the convergence rate. Recently, Müller and Wang (2016) studied the inference of extreme quantiles by what they referred to as fixed-k asymptotics. Our ABC approach takes inspiration from their idea of treating fixed-k estimates as new observations. We differ from them by considering the boundary and adopting the MCMC method for inferences. Our ABC approach is also in spirit close to the small-bandwidth asymptotics studied in Cattaneo, Crump, and Jansson (2010), because it relies on the alternative asymptotics and (approximate) finite sample inference.

The ABC method was first considered by Bickel and Yahav (1969) and Ibragimov and Has'minskii (2013). Recently, Chernozhukov and Hong (2003), Forneron and Ng (2015), Jun, Pinkse, and Wan (2015), Yu (2015), and Chen, Christensen, O'Hara, and Tamer (2016) considered ABC in M-estimations, GMM, Maximum-score type estimations, threshold regressions, and partially identified models, respectively. Creel, Gao, Hong, and Kristensen (2015) went one step further and justified the use of kernel regression instead of the MCMC method to implement the ABC approach. We contribute to this literature by applying ABC to first-stage estimates instead of the original data. We mainly exploit two advantages of the ABC method. First, it can simultaneously produce point estimates and confidence intervals. Since our parameter of interest is non-regular (i.e., not asymptotically normal), the standard inference based on normal critical values does not work. The ABC approach provides an valid alternative. Second, as has been pointed out by Hirano and Porter (2003), Chernozhukov and Hong (2004), and Ibragimov and Has'minskii (2013), the Bayesian estimator is the most efficient for non-regular cases. Our Bayesian estimator, based on the ABC method, utilizes multiple first-stage estimates in an optimal manner and automatically corrects for the mean (with L_2 loss function) and median (with L_1 loss function) bias. The idea of using ABC as an estimator-combination device appears to be new to the literature.

The rest of the article is organized as follows. Section 2 lays out the setup of the article. Section 3 derives the asymptotic properties of extreme quantiles in our context. Section 4 investigates the asymptotic properties of two inference approaches. Section 5 examines the two inference procedures on simulated data, and compares them with the procedure proposed in Daouia et al. (2010). Section 6 applies both approaches to an empirical application. We conclude with Section 7. All proofs are collected in the Appendices.

Throughout this article, capital letters, such as A, X, and Y, denote random elements while their corresponding lower cases denote realizations. C denotes an arbitrary positive constant that may not be the same in different contexts. For a sequence of random variables $\{U_n\}_{n=1}^{\infty}$ and a random variable U, $U_n \rightsquigarrow U$ indicates weak convergence in the sense of van der Vaart and Wellner (1996). Convergence in probability is denoted as $U_n \xrightarrow{p} U$.

2 Setup

Let $x \in \Re_+^p$ and $y \in \Re_+^q$ be vectors of production factors (inputs) and outputs, respectively. Technology is the set of all feasible pairs of (x, y), i.e.,

$$\mathbb{T} = \{(x, y) \in \Re_+^p \times \Re_+^q | x \text{ can produce } y\}.$$

We are interested in the estimation and inference of the production frontier (or efficient boundary) of technology $\mathbb{T}$, which is the locus of optimal production plans (maximal achievable output for a given level of inputs), i.e.,

$$\phi(x) = \sup\{y | (x, y) \in \mathbb{T}\}.$$

Researchers observe a random sample of pairs of outputs and inputs $\{X_i, Y_i\}_{i=1}^n$ such that for each $i = 1, \cdots, n$, $(X_i, Y_i) \in \mathbb{T}$.

Assumption 1. *$\{Y_i, X_i\}_{i=1}^n$ is i.i.d. $p_0 = P(X \leq x) > 0$.*

Below, we consider only the univariate output case, i.e., $q = 1$. If the outputs are multi-dimensional, then we can denote the univariate $\mathbb{Y}_i$ as the productivity efficiency score, which is defined as $\mathbb{Y}_i = \max(Y_i^1, Y_i^2, \cdots, Y_i^q)$. All the results in this article can by applied to studying the frontier of the new pair $(X_i, \mathbb{Y}_i)_{i=1}^n$. The same productivity efficiency score was also considered by Park et al. (2000) and Daouia and Simar (2007). In addition, we follow the literature and assume free disposability.

Assumption 2. *If $(x, y) \in \mathbb{T}$, then $(x', y') \in \mathbb{T}$ for any $(x'.y')$ such that $x' \geq x$ (componentwise) and $y' \leq y$.*

Let $F(y/x) = P(Y \leq y|X \leq x)$ be the "non-standard conditional distribution" in the production frontiers literature. Then under Assumption 2,

$$\phi(x) = \sup\{y \geq 0 | F(y/x) < 1\}. \tag{2.1}$$

We propose to estimate the production frontier at x by $\hat{q}_n(\tau_n)$, where

$$\hat{q}_n(\tau_n) = \arg\min_q \sum_{i=1}^n \rho_{\tau_n}(Y_i - q) \mathbb{1}\{X_i \leq x\}, \tag{2.2}$$

$\rho_\tau(u) = (\tau - \mathbb{1}\{u \leq 0\})u$ is Koenker and Bassett's (1978) check function, and τ_n is some sequence that is smaller than and converges to 1. We further denote $q(\tau_n) = F^{-1}(\tau_n/x)$ where $F^{-1}(\tau_n/x) = \inf\{y : F(y/x) \geq \tau_n\}$. We omit the dependence of $q(\tau_n)$ and $\hat{q}_n(\tau_n)$ on x for brevity. Based on this notation, $\phi(x)$, the production frontier at x, is $q(1)$.

There is a trade-off between efficiency and robustness underlying the choice of τ_n. Note that $\hat{q}_n(\tau_n)$ naturally estimates $q(\tau_n)$, and thus is a downward-biased estimator of the production frontier with bias $q(\tau_n) - q(1) \equiv q(\tau_n) - \phi(x)$. As τ_n approaches one, the bias becomes smaller. However, the estimator becomes less robust as well because a smaller proportion of the data are used for estimation. On the other hand, if τ_n is relatively too far away from one (but still close to one), the bias between $q(\tau_n)$ and $q(1)$ plays a significant role in the mean squared error (MSE), which is not asymptotically negligible. Existing inference methods of the production frontier belong to the latter case that

$$n(1-\tau_n) = k_n \to \infty \quad \text{and} \quad k_n/n \to 0, \tag{2.3}$$

where this τ_n makes the bias asymptotically non-negligible. Instead, we consider alternative asymptotics and treat τ_n as closer to one than as assumed in (2.3).

Assumption 3. $\tau_n = 1 - \frac{k}{n}$ *for some* $k \in (0, \infty)$ *and* kp_0 *is not an integer.*

The sequence of τ_n is referred to as the extreme quantile index by Chernozhukov (2005) and Daouia et al. (2010), and as fixed-k asymptotics by Müller and Wang (2016). Comparing with (2.3), the first part of Assumption 3 treats k_n as fixed at k, which does not diverge to infinite as sample size increases. However, since k can be greater than 1, we still use interior data points, rather than the maximum of the feasible outputs, for inference. Therefore, our inference procedures are robust to (a certain amount of) outliers, although it is indeed less robust than the existing inference with τ_n defined in (2.3). The second part of Assumption 3 is to guarantee that the limiting objective function of our minimization problem in (2.2) has a unique minimizer. We view this assumption as mild because we have the freedom to choose k and the integers are sparse on the real line.

3 Asymptotic Properties

Before stating the regularity condition for our asymptotic results, we first introduce some definitions. We say the cumulative distribution function (CDF) F *belongs to the domain of attraction of type III generalized extreme value (EV) distributions* if as $z \to 0$ and any $v > 0$,

$$\frac{1 - F(z_1 - vz)}{1 - F(z_1 - z)} \to v^{-1/\xi},$$

where $z_1 = \sup\{z | F(z) < 1\}$, ξ is the EV index, and $\xi < 0$.

Assumption 4. *The conditional CDF of Y_i given $X_i \le x$ belongs to the domain of attraction of type III generalized EV distributions with the EV index $\xi_0 < 0$.*

Assumption 4 states that $1 - F(y/x)$ decays polynomially as y approaching $q(1)$ or equivalently, $F(y/x)$ has a Pareto-type upper tail. This condition is common in the literature on the inference of extreme quantiles and production frontiers, e.g., Chernozhukov and Fernández-Val (2011), Daouia et al. (2010), Park et al. (2000), Zhang (2016). Appendix I contains a consistent estimator of the EV index.

Let $\alpha_n = 1/(q(1) - q(1 - 1/n))$, $\widehat{Z}_n(k) = \alpha_n(\hat{q}_n(\tau_n) - q(1))$, $\widehat{Z}_n^c(k) = \alpha_n(\hat{q}_n(\tau_n) - q(\tau_n))$, $\{\mathcal{E}_i\}_{i=1}^{\infty}$ be an i.i.d. sequence of standard exponential random variables, $\mathcal{J}_i = \eta(\sum_{l=1}^{i} \mathcal{E}_l / p_0)$ where $\eta(\cdot) = (\cdot)^{-\xi_0}$.

Theorem 3.1. *If Assumptions 1–4 hold, then $\widehat{Z}_n(k) \rightsquigarrow Z_\infty(k)$ where $Z_\infty(k) = -\mathcal{J}_h$ for some $h \in [kp_0, kp_0 + 1]$. In addition, $\widehat{Z}_n^c(k) \rightsquigarrow Z_\infty(k) + \eta(k)$.*

Several comments are in order. First, Theorem 3.1 establishes the asymptotic distribution of $\hat{q}_n(\tau_n)$, in which τ_n is of the extreme order. Second, because kp_0 is not an integer, there is exactly one integer in $[kp_0, kp_0 + 1]$. Third, Theorem 3.1 does not directly lead to a feasible inference because the convergence rate α_n is unknown. Park et al. (2000) and Simar and Wilson (2011) imposed additional assumptions so that the convergence rate becomes known up to a constant. Such conditions are not required for our approaches. The following is one way to estimate α_n from the literature. By means of Assumption 4, we know $\alpha_n = c_n n^{\xi_0}$ where ξ_0 is the EV index of the upper tail of $F(y/x)$ and c_n is a slowly varying function. In order to obtain a valid estimator of the convergence rate α_n, it is common practice to assume $c_n \to c \in (0, \infty)$. Then, a valid estimator of α_n can be constructed by replacing c and ξ_0 with their estimates. However, the assumption that $c_n \to c \in (0, \infty)$ excludes the cases where c_n decays to zero (e.g., $c_n = \log^{-1}(n)$) or diverges to infinity (e.g., $c_n = \log(n)$). In Corollary 3.1 below, we follow the idea of Bertail et al. (2004) and Chernozhukov and Fernández-Val (2011) which propose a feasible convergence rate $\hat{\alpha}_n$ that does not require any additional assumption on the tail distribution of the feasible output.

Assumption 5. *Choose two constants $k_0 > 0$ and $m > 1$ such that neither $k_0 p_0$ nor mk_0p_0 is an integer and $mk_0p_0 > k_0p_0 + 1$.*

Assumption 6. *$\hat{\omega}_1$ and $\hat{\omega}_2$ are two random weights such that $\hat{\omega}_1 + \hat{\omega}_2 = 1$, $\hat{\omega}_1 \xrightarrow{p} \omega_1$, and $\hat{\omega}_2 \xrightarrow{p} \omega_2$.*

For a generic k that satisfies Assumption 3, let $\tilde{Z}_\infty(k) = Z_\infty(k)/(Z_\infty(k_0) - Z_\infty(mk_0))$, where $Z_\infty(k) = -(\sum_{i=1}^{h(k)} \mathcal{E}_i/p_0)^{-\xi_0}$ and $h(k)$ is the unique integer that satisfies $kp_0 \leq h(k) \leq kp_0 + 1$.

Corollary 3.1. *Let $\hat{\alpha}_n = (\hat{q}_n(1-k_0/n) - \hat{q}_n(1-mk_0/n))^{-1}$, $\tau_{nl} = 1 - k_l/n$ for $l = 1, \cdots, L$. If Assumptions 1, 2, and 4–6 hold, and Assumption 3 holds for $k = k_0, mk_0, k_1, \cdots, k_L$, then*

$$\begin{pmatrix} \hat{\alpha}_n(\hat{q}_n(\tau_{n1}) - q(1)) \\ \vdots \\ \hat{\alpha}_n(\hat{q}_n(\tau_{nL}) - q(1)) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \tilde{Z}_\infty(k_1) \\ \vdots \\ \tilde{Z}_\infty(k_L) \end{pmatrix},$$

$$\hat{\alpha}_n \Big[\hat{\omega}_1 \hat{q}_n(\tau_{n1}) + \hat{\omega}_2 \hat{q}_n(\tau_{n2}) - (\omega_1 q(\tau_{n1}) + \omega_2 q(\tau_{n2})) \Big] \rightsquigarrow \frac{\omega_1 Z_\infty^c(k_1) + \omega_2 Z_\infty^c(k_2)}{Z_\infty(k_0) - Z_\infty(mk_0)}, \tag{3.1}$$

and

$$\hat{\alpha}_n\Big[\hat{\omega}_1\hat{q}_n(\tau_{n1}) + \hat{\omega}_2\hat{q}_n(\tau_{n2}) - q(1)\Big] \rightsquigarrow \frac{\omega_1 Z^c_\infty(k_1) + \omega_2 Z^c_\infty(k_2) - \omega_1\eta(k_1) - \omega_2\eta(k_2)}{Z_\infty(k_0) - Z_\infty(mk_0)}. \qquad (3.2)$$

Corollary 3.1 shows that $\hat{\alpha}_n$ is a feasible normalizing factor. The asymptotic distributions with the new normalizing factor serve as the cornerstone of our estimation and inference procedures, which we will turn to next.

4 The Estimation and Inference

As pointed out by Bickel and Freedman (1981) and Zarepour and Knight (1999), under the extreme quantile asymptotics, the standard bootstrap inference is inconsistent. In the following, we consider two alternative simulation-based inference methods.

4.1 The Subsampling Approach

We construct valid point estimators and confidence intervals for $q(1)$ based on the following two observations from Corollary 3.1. First, for $\tau_{nj} = 1 - k_j/n$, $j = 1, 2$, we can estimate $q(1)$ by $\hat{\omega}_1\hat{q}_n(\tau_{n1}) + \hat{\omega}_2\hat{q}_n(\tau_{n2})$ and the critical values of the limiting distribution of

$$\hat{\alpha}_n\Big[\hat{\omega}_1\hat{q}_n(\tau_{n1}) + \hat{\omega}_2\hat{q}_n(\tau_{n2}) - q(1)\Big]$$

is the same as those for

$$\hat{\alpha}_n\Big[\hat{\omega}_1\hat{q}_n(\tau_{n1}) + \hat{\omega}_2\hat{q}_n(\tau_{n2}) - (\omega_1 q(\tau_{n1}) + \omega_2 q(\tau_{n2}))\Big],$$

given

$$\omega_1 + \omega_2 = 1 \text{ and } \ \omega_1 \eta(k_1) + \omega_2 \eta(k_2) = 0.^{1} \tag{4.1}$$

Second, using the subsampling method, we are able to compute the critical value of the limiting distribution of

$$\hat{\alpha}_n \Big[\hat{\omega}_1 \hat{q}_n(\tau_{n1}) + \hat{\omega}_2 \hat{q}_n(\tau_{n2}) - (\omega_1 q(\tau_{n1}) + \omega_2 q(\tau_{n2})) \Big].$$

To exploit these two observations, we choose $\hat{\omega}_1$ and $\hat{\omega}_2$ that solve the sample version of (4.1):

$$\hat{\omega}_1 + \hat{\omega}_2 = 1 \quad \text{and} \quad \hat{\omega}_1 k_1^{-\hat{\xi}} + \hat{\omega}_2 k_2^{-\hat{\xi}} = 0, \tag{4.2}$$

where $\hat{\xi}$ is a consistent estimator of ξ_0. Then, we compute the critical values for the distribution of

$$(Z_\infty(k_0) - Z_\infty(mk_0))^{-1} \Big[\omega_1 Z_\infty^c(k_1) + \omega_2 Z_\infty^c(k_2) \Big],$$

which is the limiting distribution of

$$\hat{\alpha}_n \Big[\hat{\omega}_1 \hat{q}_n(\tau_{n1}) + \hat{\omega}_2 \hat{q}_n(\tau_{n2}) - (\omega_1 q(\tau_{n1}) + \omega_2 q(\tau_{n2})) \Big].$$

Using these critical values, we can construct valid estimators and confidence intervals for $q(1)$.

To implement, we compute the critical value using the subsampling method with replacement. Let b be the subsample size. For notation, the estimator computed using (2.2) with quantile index τ and the full sample is denoted as $\hat{q}_n(\tau)$. The same estimator, but computed using the subsample, is denoted as $\hat{q}_b(\tau)$. We follow the procedure below to compute the

[1] The intuition of how this linear combination reduces the downward bias is as follows. Equation (4.1) leads to a ω_1 greater than 1 and a negative $\omega_2 = 1 - \omega_1$, assuming $k_1 > k_2$. Consequently, we compensate the downward bias of $\hat{q}_n(\tau_{n1})$ to $q(1)$ by $(\omega_1 - 1)(\hat{q}_n(\tau_{n1}) - \hat{q}_n(\tau_{n2}))$.

critical value of

$$(Z_\infty(k_0) - Z_\infty(mk_0))^{-1}\left[\omega_1 Z^c_\infty(k_1) + \omega_2 Z^c_\infty(k_2)\right].$$

1. Compute $\hat{\omega}_1$ and $\hat{\omega}_2$ using (4.2) with the full sample.

2. Compute $\hat{q}_n(\tau_{n1})$, $\hat{q}_n(\tau_{n2})$, and $\hat{\alpha}_n = (\hat{q}_n(1-k_0/n) - \hat{q}_n(1-mk_0/n))^{-1}$ using the full sample. Let $\tau_{b1} = 1 - k_1/b$ and $\tau_{b2} = 1 - k_2/b$. Compute $\hat{q}_n(\tau_{b1})$ and $\hat{q}_n(\tau_{b2})$ using the full sample.

3. For the s-th subsample, compute $\hat{q}_b(\tau_{b1})$, $\hat{q}_b(\tau_{b2})$, and $\hat{\alpha}_b$, in which

$$\hat{\alpha}_b = 1/(\hat{q}_b(1-k_0/b) - \hat{q}_b(1-mk_0/b)).$$

   Define $\tilde{Z}^*_{b,s} = \hat{\alpha}_b\left(\hat{\omega}_1(\hat{q}_b(\tau_{b1}) - \hat{q}_n(\tau_{b1})) + \hat{\omega}_2(\hat{q}_b(\tau_{b2}) - \hat{q}_n(\tau_{b2}))\right)$.

4. Repeat step 3 for $s = 1, \cdots, S$ and obtain a collection of $\{\tilde{Z}^*_{b,s}\}_{s=1}^S$.

5. Denote $\hat{C}_{1-\alpha}$ as the $1-\alpha$ quantile of $\{Z^*_{b,s}\}_{s=1}^S$. Compute the median-unbiased estimator $\hat{q}^{SS}$ and $1-\alpha$ confidence interval ($\mathrm{CI}^{SS}_{1-\alpha}$) for $\omega_1 q(\tau_{n1}) + \omega_2 q(\tau_{n2})$ as

$$\hat{q}^{SS} = \hat{\omega}_1\hat{q}_n(\tau_{n1}) + \hat{\omega}_2\hat{q}_n(\tau_{n2}) - \hat{C}_{0.5}/\hat{\alpha}_n$$

   and

$$(\hat{\omega}_1\hat{q}_n(k_1) + \hat{\omega}_2\hat{q}_n(\tau_{n2}) - \hat{C}_{1-\alpha/2}/\hat{\alpha}_n, \hat{\omega}_1\hat{q}_n(\tau_{n1}) + \hat{\omega}_2\hat{q}_n(\tau_{n2}) - \hat{C}_{\alpha/2}/\hat{\alpha}_n),$$

   respectively.

Due to the non-regularity of the parameter of interest, like other inference procedures (e.g., Simar and Wilson (2011) and Daouia et al. (2010)), the subsampling procedure requires several tuning parameters, namely $\{k_0, k_1, k_2\}$, m, S, and b. We discuss these tuning parameters in detail in Section 5.3.

Assumption 7. *Assume $\hat{\xi}$ is a consistent estimator ξ_0, and $\hat{\omega}_1$ and $\hat{\omega}_2$ are computed based on* (4.2).

Unlike the asymptotic distribution, the consistency of an estimator of the EV index can be established under mild conditions, e.g., Resnick (2007). A consistent estimator of ξ_0 is contained in Appendix I.

Theorem 4.1. *If Assumptions 1, 2, and 4–7 hold, and Assumption 3 holds for $k = k_0$, mk_0, k_1 and k_2, $S \to \infty$, $b \to \infty$, and $b/n \to 0$ polynomially in n, then*

$$P(q(1) \leq \hat{q}^{SS}) \to 0.5 \quad \textit{and} \quad P(q(1) \in CI^{SS}_{1-\alpha}) \to 1 - \alpha.$$

Theorem 4.1 shows we can linearly combine two extreme quantile estimators to cancel the bias and construct a median unbiased estimator and a valid confidence interval for $q(1)$.

4.2 The ABC Approach

In this section, we consider how to combine more than two estimators in some optimal (and potentially nonlinear) manner to infer the production frontier.

Denote $\tilde{Z}_n(k_l) = \hat{\alpha}_n(\hat{q}_n(\tau_{nl}) - q(1))$ for $\tau_{nl} = 1 - k_l/n$, $l = 1, \cdots, L$. Then, Corollary 3.1 shows

$$\begin{pmatrix} \tilde{Z}_n(k_1) \\ \vdots \\ \tilde{Z}_n(k_L) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \tilde{Z}_\infty(k_1) \\ \vdots \\ \tilde{Z}_\infty(k_L) \end{pmatrix}.$$

We view $(\tilde{Z}_n(k_1), \cdots, \tilde{Z}_n(k_L))$ as new observations, whose joint density is parametrized by $q(1)$ and converges to the joint PDF of $(\tilde{Z}_\infty(k_1), \cdots, \tilde{Z}_\infty(k_L))$, which is denoted as $f(\cdot; \xi_0, p_0)$. Note the limiting density also depends on ξ_0 and p_0, because for any $l = 1, \cdots, L$, $\tilde{Z}_\infty(k_l) =$

$Z_\infty(k_l)/(Z_\infty(k_0) - Z_\infty(mk_0))$ in which $Z_\infty(k) = -(\sum_{i=1}^{h(k)} \mathcal{E}_i/p_0)^{-\xi_0}$.

Although we cannot calculate the exact finite sample likelihood of $(\tilde{Z}_n(k_1), \cdots, \tilde{Z}_n(k_L))$, we can approximate it by its limit. Then, by putting a prior on $q(1)$, we can write down the posterior distribution and conduct Bayesian inference.

To implement, we first estimate (ξ_0, p_0) by $(\hat{\xi}, \hat{p})$ so that only $q(1)$ is left unknown. Let $\pi(\cdot)$ and $\rho(\cdot)$ be the prior of $q(1)$ and a loss function, respectively. The Bayesian estimator $\hat{q}^{BE}$ of $q(1)$ minimizes the average risk, i.e.,

$$\hat{q}^{BE} = \arg\min_q \int_U \rho_n(q-\overline{q}) \frac{f(\hat{\alpha}_n(\hat{q}_n(\tau_{n1}) - \overline{q}), \cdots, \hat{\alpha}_n(\hat{q}_n(\tau_{nL}) - \overline{q}); \hat{\xi}, \hat{p})\pi(\overline{q})}{\int_U f(\hat{\alpha}_n(\hat{q}_n(\tau_{n1}) - q'), \cdots, \hat{\alpha}_n(\hat{q}_n(\tau_{nL}) - q'); \hat{\xi}, \hat{p})\pi(q')dq'} d\overline{q}, \quad (4.3)$$

where $\rho_n(u) = \rho(\hat{\alpha}_n u)$ and U is the support of $\pi(\cdot)$ that has $q(1)$ as its interior point. Let $v = \hat{\alpha}_n(\overline{q} - q(1))$, $v' = \hat{\alpha}_n(q' - q(1))$, $z = \hat{\alpha}_n(q - q(1))$, and $\hat{Z}_n^{BE} = \hat{\alpha}_n(\hat{q}^{BE} - q(1))$. Then

$$\hat{Z}_n^{BE} = \theta_n^{BE}(\tilde{Z}_n(k_1), \cdots, \tilde{Z}_n(k_L); \hat{\xi}, \hat{p}),$$

where

$$\theta_n^{BE}(z_1, \cdots, z_L; \xi, p) = \arg\min_z Q_n(z, z_1, \cdots, z_L; \xi, p), \quad (4.4)$$

$$Q_n(z, z_1, \cdots, z_L, \xi, p) = \int_{U_n} \rho(z-v) \frac{f(z_1 - v, \cdots, z_L - v; \xi, p)\pi(q(1) + v/\hat{\alpha}_n)}{\int_{U_n} f(z_1 - v', \cdots, z_L - v'; \xi, p)\pi(q(1) + v'/\hat{\alpha}_n)dv'} dv,$$

and $U_n = \hat{\alpha}_n(U - q(1))$. As $\alpha_n \to \infty$, the RHS of the above converges to

$$Q_\infty(z, z_1, \cdots, z_L; \xi, p) = \int_{\Re} \rho(z-v) \frac{f(z_1 - v, \cdots, z_L - v; \xi, p)}{\int_{\Re} f(z_1 - v', \cdots, z_L - v'; \xi, p)dv'} dv. \quad (4.5)$$

Further denote $Z_\infty^{BE} = \theta_\infty^{BE}(\tilde{Z}_\infty(k_1), \cdots, \tilde{Z}_\infty(k_L); \xi_0, p_0)$,

$$\theta_\infty^{BE}(z_1, \cdots, z_L; \xi, p) = \arg\min_z Q_\infty(z, z_1, \cdots, z_L; \xi, p), \quad (4.6)$$

and

$$\tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \xi, p) = \arg\min_{\gamma} \int_{K_t} \rho(\gamma - v) f(z_1 - v, \cdots, z_L - v; \xi, p) dv, \qquad (4.7)$$

where $K_t = [-t, t]$ for $t \geq 1$.

Assumption 8. *1. $\rho(u)$ is convex and bounded by a polynomial function of u.*

2. $(h(k_0), h(mk_0), h(k_1), \cdots, h(k_L))$ are distinct from each other.

3. $(\hat{\xi}, \hat{p}) \xrightarrow{p} (\xi_0, p_0)$.

4. $\theta_\infty^{BE}(z_1, \cdots, z_L; \xi, p)$ and $\theta_n^{BE}(z_1, \cdots, z_L; \xi, p)$ are continuous in $(z_1, \cdots, z_L) \in \Re^L$ and (ξ, p) at (ξ_0, p_0). There exist absolute constants C and B independent of n and any (ξ, p) in a neighborhood of (ξ_0, p_0), such that

$$|\theta_n^{BE}(z_1, \cdots, z_L; \xi, p)| + |\theta_\infty^{BE}(z_1, \cdots, z_L; \xi, p)| \leq C \sum_{l=1}^{L} |z_l^B| \quad a.s.$$

5. There exist constants C_t and B_t potentially dependent on t, such that, uniformly over (ξ, p) in a neighborhood of (ξ_0, p_0),

$$|\tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \xi, p)| \leq C_t \sum_{l=1}^{L} |z_l^{B_t}| \quad a.s.$$

6. $f(z_1, \cdots, z_L; \xi, p)$ is continuous in $(z_1, \cdots, z_L) \in \Re^L$ and (ξ, p) at (ξ_0, p_0), and decays exponentially to zero as $z_l \to \infty$, for all $l = 1, \cdots, L$, uniformly over a neighborhood of (ξ_0, p_0).

7. $\pi(\cdot)$ is bounded and continuous at $q(1)$.

8. $Q_\infty(z, \tilde{Z}_\infty(k_1), \cdots, \tilde{Z}_\infty(k_L); \xi_0, p_0)$ is finite over a nonempty open set $\mathcal{Z}_0$ and uniquely minimized at some random variable Z_∞^{BE} w.p.1..

Several comments are in order. First, Assumption 8.1 is common in Bayesian estimations, e.g., Chernozhukov and Hong (2003) and Chernozhukov and Hong (2004). Both l_1 and l_2 loss functions satisfy this assumption. Second, Assumption 8.2 ensures the limiting likelihood is well-defined. Third, we adopt a consistent estimator $\hat{\xi}$ of ξ_0 in Appendix I and use $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}$, which is consistent to p_0. Fourth, Assumptions 8.4 and 8.5 can be verified directly because it is possible to write down $f(z_1, \cdots, z_L; \xi, p)$ analytically. We provide one example in Proposition 4.1. Note that, unlike the standard Bayesian estimation, here we only deal with a finite sample with L observations. In the example with $L = 1$ after Theorem 4.2,

$$\theta_{\infty}^{BE}(z; \xi, p) = z - c(\xi, p)$$

in which the $c(\xi, p)$'s under l_1 and l_2 loss functions are just the median and mean of the random variable with density $f(z; \xi, p)$, respectively. If we use the uninformative prior, $\theta_n^{BE}(z_1, \cdots, z_L; \xi, p)$ is the same as $\theta_{\infty}^{BE}(z_1, \cdots, z_L; \xi, p)$, and thus Assumption 8.4 holds. Fifth, Assumption 8.5 is mild as we allow the constants C and B to depend on t. Similarly, in the example with $L = 1$,

$$\tilde{\theta}_t^{BE}(z; \xi, p) = z - c_t(z, \xi, p)$$

in which the $c_t(z, \xi, p)$'s under l_1 and l_2 loss functions are $\mathrm{med}(U|U \in z - K_t)$ and $\mathbb{E}(U|U \in z - K_t)$, respectively, where the random variable U has density $f(z; \xi, p)$. Clearly $|\tilde{\theta}_t^{BE}(z; \xi, p)| \leq C_t = t$. Sixth, Assumption 8.6 holds because $\tilde{Z}_{\infty}(k)$ behaves as a gamma random variable, whose tail decays exponentially. Seventh, Assumptions 8.1 and 8.4-8.6 induce various integrability conditions which are necessary for applying the dominated convergence theorem. Last, Assumption 8.8 implies the limiting objective function has a unique minimizer, which is necessary for applying the argmin theorem in van der Vaart and Wellner (1996). This type of assumption is common in the literature of Laplace-type estimations, e.g., Chernozhukov and Hong (2003) and Chernozhukov and Hong (2004).

Theorem 4.2. *If Assumptions 1, 2, 4, 5, and 8 hold, and Assumption 3 holds for* $k = k_0, mk_0, k_1, \cdots, k_L$*, then* $\hat{Z}_n^{BE} \rightsquigarrow Z_\infty^{BE}$.

We take the special case of $L = 1$ to illustrate the distribution of Z_∞^{BE}. When the loss function is quadratic, i.e., $\rho(u) = u^2$, Z_∞^{BE} minimizes

$$\int (z - v)^2 f(\tilde{Z}_\infty(k) - v; \xi_0, p_0) dv.$$

By the first-order condition and simple calculations, we obtain

$$Z_\infty^{BE} = \tilde{Z}_\infty(k) - \mathbb{E}\tilde{Z}_\infty(k).$$

The new limit Z_∞^{BE} is the demeaned version of the limit (i.e., $\tilde{Z}_\infty(k)$) of the original estimator. Since Z_∞^{BE} has the smallest MSE, it must be mean-unbiased. This illustrates our ABC approach can automatically correct for the bias of the original estimator. Similarly, when $\rho(u) = |u|$, the Bayesian estimator is asymptotically median-unbiased, i.e., it minimizes the mean absolute deviation (MAD).

Next, we confirm this optimality of the Bayesian estimator for the general case with $L > 1$. Let $\theta_n(\cdot; \hat{\xi}, \hat{p})$ be a (random) function of $(z_1, \cdots, z_L)$ and K be a compact subset of $\Re$. Denote the finite average risk of θ_n in K as

$$AR_{\rho,K}(\theta_n) = \int_K \int_{\Re^L} \rho(\theta_n(z_1, \cdots, z_L; \hat{\xi}, \hat{p}) - v) f(z_1 - v, \cdots, z_L - v; \hat{\xi}, \hat{p}) dz_1 \cdots d_{z_L} dv / \Lambda(K), \tag{4.8}$$

where $\rho(\cdot)$ and $\Lambda(\cdot)$ are the loss function and the Lebesgue measure, respectively. Because we treat the normalized first stage estimates $(z_1, \cdots, z_L)$ as data, $\theta_n(\cdot; \hat{\xi}, \hat{p})$, as a function of data, is also called an estimator. For a generic sequence of estimators $\{\theta_n(\cdot; \hat{\xi}, \hat{p})\}_{n\geq 1}$, the asymptotic average risk is defined as

$$AAR_\rho(\{\theta_n\}) = \limsup_{t\to\infty} \limsup_{n\to\infty} AR_{\rho,K_t}(\theta_n),$$

in which K_t is defined after (4.7).

Theorem 4.3. *If the assumptions in Theorem 4.2 hold, then*

$$AAR_\rho(\{\theta_n^{BE}\}) = \mathbb{E}\rho(Z_\infty^{BE}).$$

In addition, Let Θ_n be the collection of all estimators based on $(\tilde{Z}_n(k_1), \cdots, \tilde{Z}_n(k_L), \hat{\xi}, \hat{p})$. Then

$$\inf_{\theta_n\in\Theta_n} AAR_\rho(\{\theta_n\}) = AAR_\rho(\{\theta_n^{BE}\}).$$

Theorem 4.3 shows that the Bayesian estimator achieves the infimum of the asymptotic average risk over all estimators in Θ_n. As a corollary, by choosing the loss function to be a variant of the check function, we can show that the posterior quantiles can be used to construct valid point estimators and confidence intervals.

Corollary 4.1. *Let $\hat{q}^{BE}(0.5)$, $\hat{q}^{BE}(\tau')$, and $\hat{q}^{BE}(\tau'')$ be the Bayesian estimators that solve (4.3) with the loss function $\tilde{\rho}_\tau(u) = (\mathbb{1}\{u > 0\} - \tau)u$ and $\tau = 0.5$, τ' and τ'', respectively. Let $Z_\infty^{BE}(0.5)$, $Z_\infty^{BE}(\tau')$ and $Z_\infty^{BE}(\tau'')$ be the limits of $\hat{\alpha}_n(\hat{q}^{BE}(0.5) - q(1))$, $\hat{\alpha}_n(\hat{q}^{BE}(\tau') - q(1))$ and $\hat{\alpha}_n(\hat{q}^{BE}(\tau'') - q(1))$, respectively. If $0 < \tau' < \tau'' < 1$ and $Z_\infty^{BE}(0.5)$, $Z_\infty^{BE}(\tau')$ and $Z_\infty^{BE}(\tau'')$ are continuously distributed at zero, then*

$$P(q(1) \le \hat{q}^{BE}(0.5)) \to 0.5 \quad \text{and} \quad P(q(1) \in CI^{BE}(\tau'' - \tau')) \to \tau'' - \tau',$$

where $CI^{BE}(\tau'' - \tau') = \{\hat{q}^{BE}(\tau'), \hat{q}^{BE}(\tau'')\}$.

The Bayesian estimator $\hat{q}^{BE}(\tau)$ is just the τ-th posterior quantile. Corollary 4.1 shows we can construct a median-unbiased estimator and a valid confidence interval based on posterior quantiles. To implement the MCMC method (such as the Metropolis-Hastings algorithm) and obtain the posterior distribution, we have to evaluate $f(\cdot;\xi,p)$ at

$$(\hat{\alpha}_n(\hat{q}_n(\tau_{n1})-\overline{q}),\cdots,\hat{\alpha}_n(\hat{q}_n(\tau_{nL})-\overline{q})).$$

Next, we derive an analytical form of $f(u_1,\cdots,u_L;\xi,p)$, which is the joint PDF of

$$(\tilde{Z}_\infty(k_1),\cdots,\tilde{Z}_\infty(k_L)).$$

Assumption 9. $h(k_0)<h(mk_0)<h(k_1)<\cdots<h(k_L)$.

The order of h's is needed to derive a simple formula for the joint PDF but is not required for Theorem 4.2. Essentially, Assumption 9 requires that $h(k_0)$ and $h(mk_0)$ are smaller than all the other h's, which makes it much easier to handle the common denominator $Z_\infty(mk_0)-Z_\infty(k_0)$ in $\tilde{Z}_\infty(k_l)$ for $l=1,\cdots,L$.

Proposition 4.1. *Let f_h be the PDF of a gamma random variable with shape and scale parameters being equal to h and 1, respectively. If Assumption 9 holds, then*

$$\begin{aligned}&f(u_1,\cdots,u_L;\xi,p)\\&=\int(-1/\xi)^L\tilde{u}(t,s)^{-L/\xi}\left[\prod_{l=1}^{L}u_l^{-1/\xi-1}f_{h_l-h_{l-1}}(v_l-v_{l-1})\right]f_{h(k_0)}(s)f_{h(mk_0)-h(k_0)}(t)dsdt,\end{aligned}$$

where $h_l=h(k_l)$ for $1\le l\le L$, $h_0=h(mk_0)$, $v_l=(u_l\tilde{u}(t,s))^{-1/\xi}$ for $1\le l\le L$, $\tilde{u}(t,s)=(t+s)^{-\xi}-s^{-\xi}$, and $v_0=t+s$.

Using Proposition 4.1, for any given (s, t), we can analytically compute

$$(-1/\xi)^L \tilde{u}(t,s)^{-L/\xi} \left[\prod_{l=1}^{L} u_l^{-1/\xi - 1} f_{h_l - h_{l-1}}(v_l - v_{l-1}) \right].$$

We then compute $f(u_1, \cdots, u_L; \xi, p)$ by generating (s, t) independently as gamma random variables with parameters $(h(k_0), 1)$ and $(h(mk_0) - h(k_0), 1)$, respectively, and applying the Monte Carlo integration. Given the analytical form of $f(u_1, \cdots, u_L; \xi, p)$, the estimates $(\hat{q}_n(\tau_{n1}), \cdots, \hat{q}_n(\tau_{nL}))$, and the feasible convergence rate $\hat{\alpha}_n$, we can generate MCMC draws from the posterior

$$f(\hat{\alpha}_n(\hat{q}_n(\tau_{n1}) - \overline{q}), \cdots, \hat{\alpha}_n(\hat{q}_n(\tau_{nL}) - \overline{q}); \hat{\xi}, \hat{p})\pi(\overline{q}).$$

Then, we can use these MCMC draws to compute $\hat{q}^{BE}$, which is just the mean or median of the posterior sample depending on whether $\rho(u) = u^2$ or $\rho(u) = |u|$ is used. We can also construct $\mathrm{CI}^{BE}(1-\alpha)$ using the $\alpha/2$-th and $(1-\alpha/2)$-th posterior quantiles. Similar to the subsampling approach, the ABC approach also requires several tuning parameters, namely $(k_0, \cdots, k_L)$ and m. We will discuss the choices of these tuning parameters in Section 5.3.

5 Simulations

In this section, we investigate the finite-sample performances of our estimation and inference procedures.

5.1 Data generating processes

We consider the following two data generating processes (DGPs), which have been considered in various previous papers, e.g., Aragon et al. (2005), Martins-Filho and Yao (2008), and

Daouia et al. (2010):

DGP 1: $Y_i = X_i^{0.5}\mathcal{U}_i$, $i = 1, 2, \ldots, n$, $X \sim \text{Unif}(0, 6)$ and $\mathcal{U} \sim \text{Unif}(0, 1)$,

DGP 2: (X_i, Y_i), $i = 1, 2, \ldots, n$, are uniformly distributed in the triangle $\{(x, y) : 0 \leq x \leq 6, 0 \leq y \leq x\}$.

We assume $\{X_i, Y_i\}_{i=1}^n$ in both DGPs are i.i.d. sequences. The frontier $\phi(x)$ is $x^{0.5}$ and x for the first and second DGP, respectively. Through simple calculations, the EV index $\xi_0 = -0.5$ for all $x \in [0, 6]$ in both settings. In addition, $P(X \leq x) = x/6$ and $P(X \leq x) = (x/6)^2$ for $0 \leq x \leq 6$ in DGP 1 and 2, respectively.

5.2 Estimation of the EV index

We compare three different inference procedures, namely, the two procedures we propose in Section 4, and the one based on the intermediate quantile of feasible output, as proposed by Daouia et al. (2010) (i.e., $\hat{\varphi}_1^*(x)$ in their paper). All three inference procedures require the estimation of the EV index ξ_0. We obtain $\hat{\xi}$ via the standard Pickands-type estimator as discussed in Appendix I, i.e.,

$$\hat{\xi} = \frac{1}{\log(2)} \log\left(\frac{\hat{q}_n(4\tau_n) - \hat{q}_n(2\tau_n)}{\hat{q}_n(2\tau_n) - \hat{q}_n(\tau_n)}\right).$$

Resnick (2007) and Daouia et al. (2010) have suggested using a τ_n in $(0, 0.25]$. Since the estimation of ξ_0 is not the main focus of this article, we simply set $\tau_n = 0.1$ and $\tau_n = 0.08$ for $n = 5,000$ and $n = 10,000$, respectively, following Zhang (2016) and Chan, Hou, and Yang (2017). Daouia et al. (2010, Section 3.2) proposed a sophisticated data-driven procedure to select τ_n. Some preliminary simulations show that both the simple method and the data-driven method of choosing τ_n work well and the corresponding results are similar. In addition, in order to ensure a fair comparison, for each Monte Carlo replication, all three inference procedures considered are forced to use the same $\hat{\xi}$.

5.3 Tuning parameters

Other than the tuning parameter τ_n in the estimation of ξ_0, the remaining tuning parameters are the spacing parameter m, the extreme quantile indices $\{k_l\}_{l=0}^{L}$,[2] the number of quantiles used in the ABC method L, the number of subsamples S, and the subsample size b. How to choose those tuning parameters optimally is an important yet challenging problem. Here, we provide some rules of thumb based on either the existing literature on extreme quantile estimation or our own simulation experience. We leave the formal analysis on the higher-order impact of the tuning parameters to future research.

Given Theorem 3.1, the effective quantile indices that affect the asymptotic behaviors of our estimators are $\{h(k_l)\}_{l=1}^{L}$. We choose $\{h(k_l)\}_{l=1}^{L}$[3] from a range $[h_1, h_2]$. The lower bound h_1 is determined by how much tolerance we have for outliers. Based on the theoretical results in the previous section, both our inference approaches are robust to $\lfloor h_1 \rfloor$ number of outliers. The upper bound h_2 takes the form of $\max(C_1, C_2 b\hat{p})$ for two constants C_1 and C_2. First, in order to guarantee a good approximation of extreme quantile asymptotics, Chernozhukov and Fernández-Val (2011) suggested that the effective quantile indices should be less than 40 or 80. To be cautious, we choose $C_1 = 40$. Second, for the subsample with size b, the effective subsample size is bp_0. In order to guarantee a good approximation of extreme quantile asymptotics for the subsample, we need $\tau = h(k)/bp_0$ to be close to zero. This implies the second requirement that $h(k) \leq C_2 bp_0$. We view τ is sufficiently close to zero if $\tau \leq 0.1$ and set $C_2 = 0.1$. This gives the upper bound that $h_2 = \max(40, 0.1b\hat{p})$ by replacing p_0 with its estimator $\hat{p}$. Furthermore, for numerical stability, we choose $\{k_l\}_{l=0}^{L}$ and m to ensure equal spaces between those effective quantile indices to be used, that is, $(h(k_0), h(mk_0), h(k_1), \cdots, h(k_L))$. We provide more detail on the choice of m below. For the choice of b, we suggest the following rule from D'Haultfoeuille, Maurel, and Zhang (2017)

[2] For the subsampling method, only (k_0, k_1, k_2) are used even if $L \geq 3$.

[3] When computing $h(k)$, the estimator $\hat{p}$ of p is used.

and Zhang (2016) which works well for samples with moderate sizes (e.g., from 800 to 10,000):

$$b = \left\lfloor \left[0.4n\hat{p} - \frac{1}{7}(n\hat{p} - 300)_+ - \frac{2.3}{28}(n\hat{p} - 1000)_+ - \frac{7}{40}\left(1 - \frac{\log 5000}{\log n\hat{p}}\right)(n\hat{p} - 5000)_+\right] \Big/ \hat{p} \right\rfloor,$$

where $(\cdot)_+ = \max(\cdot, 0)$ and $\lfloor u \rfloor$ is the floor of u. For a fixed x, the effective sample and subsample sizes are np_0 and bp_0, respectively. The proposed b satisfies $bp_0 \to \infty$ and $\frac{bp_0}{np_0} \to 0$ as $np_0 \to \infty$. For L, we do not recommend $L > 3$ for np_0 under $5,000$. We also prefer $L = 2$ to $L = 3$ for np_0 under 2,000. Our concern is that, in order to use higher fixed quantile indices for larger L, we need larger samples to approximate them well.

For practical implementation, the number of replications S for the subsampling approach and the lengths of both the burn-in sequence and the whole MCMC sequence for the ABC approach should be set as large as computationally possible. We suggest using 5,000, 10,000, and 20,000, respectively. For the initial value of the MCMC, we use the point estimator of some high quantile computed from (2.2). The acceptance rate for the MCMC sequence should be around 30%. In our implementation, we use the Metropolis-Hastings algorithm with a Gaussian proposal distribution. The standard deviation σ of the Gaussian proposal distribution serves as a tuning parameter to control the acceptance rate. We find that setting σ to be equal to or slightly less than the difference between the 97.5% and 2.5% quantiles of the posterior distribution usually can result in about 30% acceptance rate.

We denote our subsampling approach, ABC approach with $L = 2$ and $L = 3$, and the method proposed in Daouia et al. (2010) as "Sub," "ABC $L = 2$," "ABC $L = 3$" and "DFS," respectively. For our approaches, we report the results from two sets of $(\{k_l\}_{l=0}^3, m)$, which are denoted as S1 and S2 and correspond to

$$(h(k_0), h(mk_0), h(k_1), h(k_2), h(k_3)) = (15, 21, 27, 33, 39) \quad \text{and} \quad (10, 15, 20, 25, 30),$$

respectively. Note that $m = \frac{21}{15} \approx 1.4$ and 1.5 in S1 and S2, respectively. Chernozhukov and

Fernández-Val (2011) suggested using $m = 1 + \frac{sp}{k_0}$, where $sp \in [2, 20]$.[4] In S1 and S2, the corresponding sp's are 6 and 5, respectively, which satisfy the formula. When computing "Sub" and "ABC $L = 2$," we use only $(h(k_0), h(mk_0), h(k_1), h(k_2))$. In addition, we use $h(k_3)$ when computing "ABC $L = 3$." Based on our further simulations, other sets of $(\{k_l\}_{l=0}^{3}, m)$ which satisfy the above rule of thumb also work well. We do not report them due to the length limit. Furthermore, we omit the results corresponding to "ABC $L = 3$ S2" because these results are similar to those for "ABC $L = 3$ S1." For "DFS", we report the results with two sets of tuning parameters, namely, $k_n = 0.15nP$ and $k_n = 0.20nP$ (k_n from their notations). These two sets are denoted as "DFS S1" and "DFS S2." Finally, we set the number of observations to $n = 5,000$. Further simulation results with $n = 10,000$ can be found in Appendix J. All simulations are repeated $5,000$ times.

5.4 Results

We construct 95% confidence intervals using three procedures, namely "Sub," "ABC," and "DFS." Daouia et al. (2010) found their method does not work well for effective sample size np_0 under 1,000. We find similar results for our approaches. We conjecture it is mainly because the EV index estimator is too imprecise in this case. We only report reasonable results of the coverage probabilities and average lengths of the CIs for $\phi(x)$ at $x = 2.2, 3.3, 4.4, 5.5$ and $x = 3.3, 4.4, 5.5$ for DGP 1 and 2, respectively. When $n = 5,000$, the corresponding minimum effective sample sizes np_0 are 1,833 and 1,513, respectively.

Tables 1 and 2 report the coverage rates and average lengths of the 95% confidence intervals for each inference procedure. Further improved results with $n = 10,000$ can be found in Appendix J.

[4] The original formula in Chernozhukov and Fernández-Val (2011) is $m = 1 + \frac{d+sp}{k_0}$ where d is the dimension of the regressors. In our case, there is no regressor so $d = 0$.

Table 1: DGP 1, $n = 5,000$

	Sub		ABC			DFS	
	S1	S2	L=2 S1	L=2 S2	L=3 S1	S1	S2
$x = 2.2$	0.943	0.952	0.921	0.929	0.900	0.940	0.920
$np_0 \approx 1833$	(0.502)	(0.503)	(0.236)	(0.225)	(0.192)	(0.453)	(0.453)
$x = 3.3$	0.945	0.951	0.928	0.934	0.920	0.971	0.953
$np_0 \approx 2750$	(0.486)	(0.489)	(0.233)	(0.222)	(0.190)	(0.539)	(0.539)
$x = 4.4$	0.944	0.948	0.937	0.945	0.925	0.984	0.974
$np_0 \approx 3667$	(0.478)	(0.481)	(0.232)	(0.221)	(0.189)	(0.615)	(0.615)
$x = 5.5$	0.952	0.946	0.940	0.941	0.938	0.994	0.987
$np_0 \approx 4583$	(0.476)	(0.479)	(0.232)	(0.221)	(0.188)	(0.686)	(0.686)

Notes: The coverage rates and average lengths of the CIs (in parentheses) are reported.

Table 2: DGP 2, $n = 5,000$

	Sub		ABC			DFS	
	S1	S2	L=2 S1	L=2 S2	L=3 S1	S1	S2
$x = 3.3$	0.945	0.947	0.910	0.913	0.890	0.921	0.895
$np_0 \approx 1513$	(1.247)	(1.245)	(0.577)	(0.558)	(0.469)	(1.026)	(1.026)
$x = 4.4$	0.945	0.946	0.932	0.934	0.918	0.969	0.951
$np_0 \approx 2689$	(1.193)	(1.203)	(0.571)	(0.547)	(0.462)	(1.306)	(1.307)
$x = 5.5$	0.945	0.946	0.941	0.938	0.934	0.992	0.986
$np_0 \approx 4201$	(1.159)	(1.195)	(0.573)	(0.549)	(0.465)	(1.613)	(1.614)

Notes: The coverage rates and average length of the CIs (in parentheses) are reported.

Several remarks are in order. First, comparing the coverage rates across three methods, we find the subsampling approach performs best. The ABC approach always undercovers, but its coverage rates improve and converge to the nominal 95% as the effective sample size increases. We conjecture this improvement is due to the fact that the estimator of the EV index becomes more accurate as the effective sample increases. The CIs constructed using the DFS approach undercover when the effective sample size is small and overcover when the effective sample size is large. However, it should be noted that, for simplicity, we do not choose the optimal tuning parameters (especially k_n) for DFS. This may be one of the reasons for their relatively inferior performances.

Second, comparing the average lengths across three methods, we find the ABC approach has the shortest CI. In particular, "ABC $L = 2$" and "Sub" are based on the same set of extreme quantile estimators, but the length of CIs for the latter are more than double. This confirms our theoretical results that the ABC approach is more efficient. In addition, the average lengths of "ABC $L = 3$" are about 20% shorter than those of "ABC $L = 2$." This implies there is information gain by including more extreme quantile estimators. However, the coverage rates also become worse when $L = 3$. Between the subsampling approach and DFS, the average lengths are comparable but the subsampling approach performs better when the effective sample size is greater than 2,000.

Third, comparing the results for S1 and S2, we find both the coverage rates and average lengths are quite stable, for both subsampling and ABC approaches. This indicates neither approach is sensitive to the tuning parameters, as long as the rules of thumb are satisfied.

Finally, comparing the results with $n = 5,000$ and $n = 10,000$ in Appendix J, we find both the coverage rates and the average lengths improve as the sample size increases. This indicates the validity of the fixed-k type asymptotics, which both our approaches rely on. The coverage rates of DFS become larger as we use larger samples and sometimes go up to 99%. But again, this may be because we do not choose the tuning parameters optimally for DFS.

Tables 6–9 in Appendix J also contain the performances of the point estimators from the three procedures. Overall, the posterior median from the ABC approach with $L = 3$ has the best performance in terms of mean absolute deviations and root mean squared errors.

To sum up, our methods work well with and are not sensitive to reasonable choices of tuning parameters. The subsampling approach has very accurate and stable coverage rates while the ABC approach produces the shortest CIs and most accurate point estimators. However, we also want to emphasize that these results do not mean our methods ourperform the existing method in the literature. First, the performances of "DFS" reported here can still be improved. Second, the procedure set forth in Daouia et al. (2010) is based on the intermediate, rather than extreme, quantile estimators. This indicates their approach can tolerate more outliers. As put by Daouia et al. (2010), " *It is difficult to imagine one procedure being preferable in all contexts. Hence, a sensible practise is not to restrict the frontier analysis to one procedure . . . .*" We view our approaches as alternatives to the existing inference procedures in the literature. Our simulation study shows the performances of our approaches are satisfying.

6 An Empirical Application

We apply our inference approaches to the frontier analysis of French post offices observed in 1994. The same dataset is also used in Daouia et al. (2010). In this context, X and Y denote the quantity of labor and the logarithm of volume of the delivered mail, respectively.[5] The total number of observations is 4,000, which is close to what we consider in our simulations. The summary statistics of the data are in Table 3.

Table 3: Summary Statistics

	MEAN	STD	MIN	LQ	MEDIAN	UQ	MAX
X	1592	790	177	1128	1338	1730	4405
Y	7.709	0.612	3.829	7.349	7.698	8.062	9.576

Notes: STD = standard errors, LQ = 25% quantile, UQ = 75% quantile.

We use the same sets of tuning parameters as in the simulations with details in Section 5.3. The "DFS S1" estimator of the frontier serves as the benchmark. It appears there are two possible outliers in the data (shown as circles in Figure 1). We then report the results with and without outliers in Figures 1 and 2, respectively. Given our $h(k_0)$ is greater than 2 in both setups "S1" and "S2," even with the outliers, our inference procedures should still

[5]We use log of the volume of delivered mail to smooth the data; otherwise data points are too scarce around the frontier, which makes the estimation and inference volatile. We thank Valentin Zelenyuk for his insightful suggestion on this transformation.

be valid.

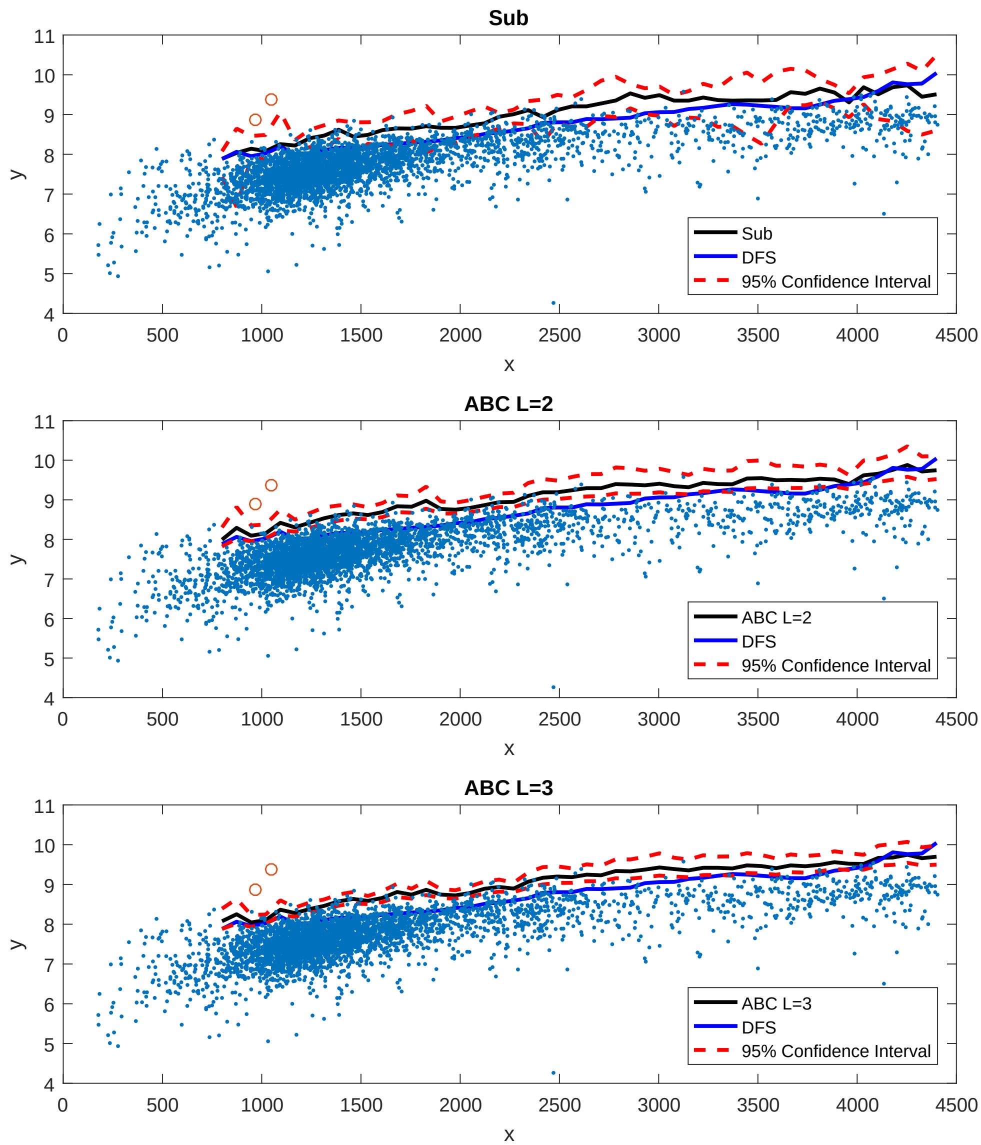

Figure 1: Estimation and Inference with Outliers

We consider three inference procedures, namely “Sub S1,” “ABC $L = 2$ S1, ”and “ABC $L = 3$ S1.” The point estimators and the associated 95% confidence intervals of the frontier at each value of $(800, 850, 900, ..., \max(X))$ are reported.

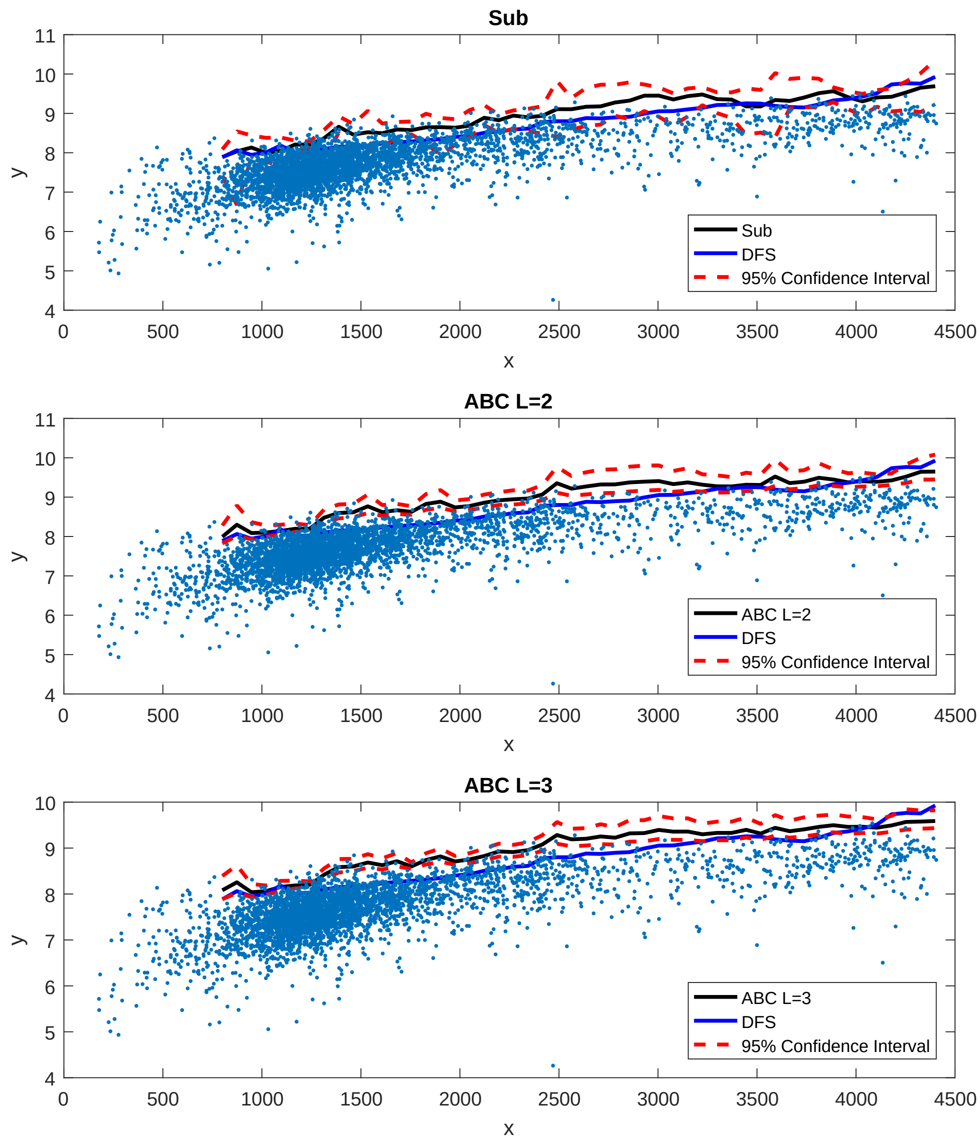

Figure 2: Estimation and Inference without Outliers

Several comments regarding Figures 1 and 2 are in order. First, the point estimators and confidence intervals from our approaches are basically the same with or without the outliers, confirming the robustness of our approaches. Second, the point estimators of our approaches are above those of "DFS S1." It is clear from the figures that our estimators in general envelope the data while "DFS S1" does not. The lengths of the confidence intervals of the ABC approach is shorter than those of the subsampling approach, which is consistent with our theoretical and simulation findings. Specifically, the average lengths of the CIs of "Sub S1" and "ABC $L = 2$ S1" over $x \in (800, 850, 900, ..., \max(X))$ are 0.731 and 0.430, respectively. The ratio of the average lengths between these two approaches is in line with the one obtained from our simulations.

7 Conclusion

In this article, we propose two novel estimation and inference procedures for production frontiers. Our procedures are based on extreme quantile estimators, and thus are robust to a few outliers. The subsampling approach has stable coverage rates across different effective sample sizes while the ABC approach is more efficient and has shorter CIs. The asymptotic validity of both procedures is theoretically justified. The application to the French post offices dataset shows that these two approaches are practical alternatives to the existing inference methods in the literature.

References

Aragon, Y., A. Daouia, and C. Thomas-Agnan (2005). Nonparametric frontier estimation: a conditional quantile-based approach. *Econometric Theory 21*(02), 358–389.

Bertail, P., C. Haefke, D. N. Politis, and H. White (2004). Subsampling the distribution of diverging statistics with applications to finance. *Journal of Econometrics 120*(2), 295–326.

Bickel, P. J. and D. A. Freedman (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics 9*(6), 1196–1217.

Bickel, P. J. and J. A. Yahav (1969). Some contributions to the asymptotic theory of Bayes solutions. *Probability Theory and Related Fields 11*(4), 257–276.

Cattaneo, M. D., R. K. Crump, and M. Jansson (2010). Robust data-driven inference for density-weighted average derivatives. *Journal of the American Statistical Association 105*(491), 1070–1083.

Cazals, C., J.-P. Florens, and L. Simar (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics 106*(1), 1–25.

Chan, J., C. Hou, and T. T. Yang (2017). Asymptotic trimming for importance sampling estimators with infinite variance. *Working paper*.

Chen, X., T. M. Christensen, K. O'Hara, and E. Tamer (2016). MCMC confidence sets for identified sets. *Working paper*.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics 33*(2), 806–839.

Chernozhukov, V. and I. Fernández-Val (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *The Review of Economic Studies 78*(2), 559–589.

Chernozhukov, V. and H. Hong (2003). An MCMC approach to classical estimation. *Journal of Econometrics 115*(2), 293–346.

Chernozhukov, V. and H. Hong (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica 72*(5), 1445–1480.

Creel, M., J. Gao, H. Hong, and D. Kristensen (2015). Bayesian indirect inference and the ABC of GMM. *Working paper*.

Daouia, A., J.-P. Florens, and L. Simar (2010). Frontier estimation and extreme value theory. *Bernoulli 16*(4), 1039–1063.

Daouia, A., S. Girard, and A. Guillou (2014). A Γ-moment approach to monotonic boundary estimation. *Journal of Econometrics 178*(2), 727–740.

Daouia, A. and L. Simar (2007). Nonparametric efficiency analysis: a multivariate conditional quantile approach. *Journal of Econometrics 140*(2), 375–400.

Deprins, D., L. Simar, and H. Tulkens (1984). Measuring labor-efficiency in post offices. In *M. Marchand, P. Pestieau and H. Tulkens (eds), The Performance of Public Enterprises: Concepts and Measurements*, pp. 243–267. Amsterdam: North-Holland.

D'Haultfoeuille, X., A. Maurel, and Y. Zhang (2017). Extremal quantile regressions for selection models and the black-white wage gap. Working Paper.

Forneron, J.-J. and S. Ng (2015). The ABC of simulation estimation with auxiliary statistics. *Working paper*.

Gattoufi, S., M. Oral, and A. Reisman (2004). Data envelopment analysis literature: A bibliography update (1951–2001). *Journal of Socio-Economic Planning Sciences 38*(2-3), 159–229.

Geyer, C. J. (1996). On the asymptotics of convex stochastic optimization. *Technical report, Dept. Statistics, Univ. Minnesota.*.

Gijbels, I., E. Mammen, B. U. Park, and L. Simar (1999). On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association 94*(445), 220–228.

Hirano, K. and J. R. Porter (2003). Asymptotic efficiency in parametric structural models with parameter-dependent support. *Econometrica 71*(5), 1307–1338.

Ibragimov, I. A. and R. Z. Has'minskii (2013). *Statistical Estimation: Asymptotic Theory*, Volume 16. Springer Science & Business Media.

Jeong, S.-O. (2004). Asymptotic distribution of DEA efficiency scores. *Journal of the Korean Statistical Society 33*, 449–458.

Jeong, S.-O. and B. U. Park (2006). Large sample approximation of the distribution for convex-hull estimators of boundaries. *Scandinavian Journal of Statistics 33*(1), 139–151.

Jun, S. J., J. Pinkse, and Y. Wan (2015). Classical Laplace estimation for $\sqrt[3]{n}$-consistent estimators: Improved convergence rates and rate-adaptive inference. *Journal of Econometrics 187*(1), 201–216.

Kneip, A., B. U. Park, and L. Simar (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory 14*(6), 783–793.

Kneip, A., L. Simar, and P. W. Wilson (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory 24*(6), 1663–1697.

Knight, K. (1999). Epi-convergence and stochastic equi-semicontinuity. *Technical report, Dept. Statistics, Univ. Toronto*.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Martins-Filho, C. and F. Yao (2008). A smooth nonparametric conditional quantile frontier estimator. *Journal of Econometrics 143*(2), 317–333.

Müller, U. K. and Y. Wang (2016). Fixed-k asymptotic inference about tail properties. *Journal of the American Statistical Association* (just-accepted).

Park, B. U., S.-O. Jeong, and L. Simar (2010). Asymptotic distribution of conical-hull estimators of directional edges. *The Annals of Statistics 38*(3), 1320–1340.

Park, B. U., L. Simar, and C. Weiner (2000). The FDH estimator for productivity efficiency scores. *Econometric Theory 16*(6), 855–877.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory 7*(2), 186–199.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes.* New York: Springer-Verlag.

Resnick, S. I. (2007). *Heavy-tail Phenomena: Probabilistic and Statistical Modeling.* Springer Science & Business Media.

Simar, L. and P. W. Wilson (2011). Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis 36*(1), 33–53.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* New York: Springer.

Yu, P. (2015). Adaptive estimation of the threshold point in threshold regression. *Journal of Econometrics 189*(1), 83–100.

Zarepour, M. and K. Knight (1999). Bootstrapping point processes with some applications. *Stochastic Processes and Their Applications 84*(1), 81–90.

Zhang, Y. (2016). Extremal quantile treatment effects. *Working paper*.

A Proof of Theorem 3.1

Denote $\mathbb{L}(u,v) = (v-u)\mathbb{1}\{u < v\}$.

$$
\begin{aligned}
\widehat{Z}_n(k) &= \arg\min_z \sum_{i=1}^n \frac{1}{\alpha_n}\Big[\alpha_n(Y_i - q(1)) - z\Big]\Big[1 - \frac{k}{n} - \mathbb{1}\{\alpha_n(Y_i - q(1)) \le z\}\Big]\mathbb{1}\{X_i \le x\} \\
&= \arg\min_z \sum_{i=1}^n \frac{1}{\alpha_n}\Big[\alpha_n(Y_i - q(1)) - z\Big]\Big[\mathbb{1}\{\alpha_n(Y_i - q(1)) > z\} - \frac{k}{n}\Big]\mathbb{1}\{X_i \le x\} \\
&= \arg\min_z \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i \le x\}kz + \sum_{i=1}^n \mathbb{L}(-\alpha_n(Y_i - q(1)), -z)\mathbb{1}\{X_i \le x\} \\
&= \arg\min_z \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i \le x\}kz + \int \mathbb{L}(u, -z)d\hat{N}_n \\
&= -\arg\min_z -\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i \le x\}kz + \int \mathbb{L}(u, z)d\hat{N}_n,
\end{aligned}
\tag{A.1}
$$

where $\hat{N}_n = \sum_{i=1}^n \mathbb{1}\{-\alpha_n(Y_i - q(1)) \in \cdot, X_i \le x\}$ is a point process. We denote

$$
-\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i \le x\}kz + \int \mathbb{L}(u, z)d\hat{N}_n
$$

as the sample objective function. We derive the asymptotic distribution of $\widehat{Z}_n(k)$ in two steps. In the first step, we derive the limit of the sample objective function point-wise in z. Since the check function $\rho_\tau(u)$ and thus the sample objective function are convex, the point-wise convergence in z is sufficient for the uniform convergence in z. Given the uniform convergence of the sample objective function, in the second step we show that the limit objective function has a unique minimizer $Z_\infty(k)$ with probability one. Then, we can apply the argmin theorem detailed in van der Vaart and Wellner (1996) to establish the weak convergence of $\widehat{Z}_n(k)$.

Step 1:

For the RHS of (A.1), the first term converges to $-p_0 kz$ point-wise in z. For the second term, we can show that the point process $\hat{N}$ weakly converges to N, a Poisson random measure with mean measure $\mu([a,b]) = p_0(\eta^{-1}(b) - \eta^{-1}(a))$. In addition, note that both $\hat{N}_n$ and N are random measures on $\Re^+ = [0, \infty)$ because $Y_i \leq q(1)$ for any $i \geq 1$. Then for any fixed $z \geq 0$ and $u \in \Re^+$, $|\mathbb{L}(u,z)|$ is bounded by z, vanishes for $u \geq z$, and is continuous in u. Then, by the continuous mapping theorem, we have, point-wise in z,

$$\int \mathbb{L}(u,z)d\hat{N}_n \rightsquigarrow \int \mathbb{L}(u,z)dN.$$

Now we show

$$\hat{N}_n \rightsquigarrow N.$$

Let $U_i = \alpha_n(Y_i - q(1))$. Given Chernozhukov (2005, Lemma 9.3 and 9.4), it suffices to show that, for any $0 \leq a < b < \infty$,

$$nP(-U_i \in [a,b], X_i \leq x) \to p_0(\eta^{-1}(b) - \eta^{-1}(a)).$$

Note that $F(y/x) \equiv P(Y \leq y | X \leq x)$. Then,

$$\begin{aligned}
nP(-U_i \in [a,b], X_i \leq x) =& np_0 P(-U_i \in [a,b] | X_i \leq x) \\
=& p_0 P(Y_i \in [q(1) - \frac{b}{\alpha_n}, q(1) - \frac{a}{\alpha_n}] | X_i \leq x)/(1 - F(q(1) - \frac{1}{\alpha_n}/x)) \\
=& p_0 \frac{F(q(1) - \frac{a}{\alpha_n}/x) - F(q(1) - \frac{b}{\alpha_n}/x)}{1 - F(q(1) - \frac{1}{\alpha_n}/x)} \\
\to& p_0(\eta^{-1}(b) - \eta^{-1}(a)).
\end{aligned}$$

By Resnick (1987, Proposition 3.7 and 3.8), $N(\cdot)$ can be written as $\sum_{i=1}^{\infty} \mathbb{1}\{\mathcal{J}_i \in \cdot\}$. Therefore, the sample objective function will converge to

$$-kp_0 z + \int \mathbb{L}(u,z)dN = -kp_0 z + \sum_{i=1}^{\infty} \mathbb{L}(\mathcal{J}_i, z)$$

weakly and uniformly over $z \in \Re^{+}$.

Step 2:

From the first-order condition of the limit objective function, we find that $Z_\infty(k) = -J_h$ for an integer h satisfying $kp_0 \le h \le kp_0 + 1$. Since kp_0 is not an integer, the limiting objective function has a unique minimizer. This concludes the proof of the first part of the theorem.

For $\widehat{Z}_n^c(k)$, we note that

$$\widehat{Z}_n^c(k) = \widehat{Z}_n(k) + \alpha_n(q(1) - q(\tau_n)) \rightsquigarrow Z_\infty(k) + \eta(k)$$

because

$$\alpha_n(q(1) - q(\tau_n)) = \frac{q(1) - q(1 - \frac{k}{n})}{q(1) - q(1 - \frac{1}{n})} \to \eta(k).$$

B Proof of Corollary 3.1

For the first claim, denote $\hat{Z}_n(k) = \alpha_n(\hat{q}_n(1 - k/n) - q(1))$, $k = k_0, mk_0, k_1, \cdots, k_L$. By the proof of Theorem 3.1, we have

$$\begin{pmatrix} \hat{Z}_n(k_0) \\ \hat{Z}_n(mk_0) \\ \hat{Z}_n(k_1) \\ \vdots \\ \hat{Z}_n(k_L) \end{pmatrix} \rightsquigarrow \begin{pmatrix} Z_\infty(k_0) \\ Z_\infty(mk_0) \\ Z_\infty(k_1) \\ \vdots \\ Z_\infty(k_L) \end{pmatrix},$$

where

$$\begin{aligned} &(-Z_\infty(k_0), -Z_\infty(mk_0), -Z_\infty(k_1), \cdots, -Z_\infty(k_L)) \\ = &\underset{z_0, z_{m0}, z_1, \cdots, z_L}{\arg\min}\; Q_\infty(z_0, k_0) + Q_\infty(z_{m0}, mk_0) + \sum_{l=1}^{L} Q_\infty(z_l, k_l) \end{aligned}$$

and

$$Q_\infty(z,k) = -kp_0 z + \sum_{i=1}^{\infty} \mathbb{L}(\mathcal{J}_i, z).$$

Then we have

$$\begin{pmatrix} \hat{\alpha}_n(\hat{q}_n(\tau_{n1}) - q(1)) \\ \vdots \\ \hat{\alpha}_n(\hat{q}_n(\tau_{nL}) - q(1)) \end{pmatrix} \rightsquigarrow \begin{pmatrix} Z_\infty(k_1)/(Z_\infty(k_0) - Z_\infty(mk_0)) \\ \vdots \\ Z_\infty(k_L)/(Z_\infty(k_0) - Z_\infty(mk_0)) \end{pmatrix} = \begin{pmatrix} \tilde{Z}_\infty(k_1) \\ \vdots \\ \tilde{Z}_\infty(k_L) \end{pmatrix}.$$

The denominator $Z_\infty(k_0) - Z_\infty(mk_0)$ is nonzero because $Z_\infty(k_0) = -\mathcal{J}_{h_0}$ and $Z_\infty(mk_0) = -\mathcal{J}_{h_0'}$ for $h_0 \in (k_0p_0, k_0p_0+1)$ and $h_0' \in (mk_0p_0, mk_0p_0+1)$, respectively, and by Assumption 5, $h_0 \neq h_0'$ because $mk_0p_0 > k_0p_0 + 1$.

For the second result, note that

$$\hat{\omega}_1\hat{\alpha}_n(\hat{q}_n(\tau_{n1}) - q(\tau_{n1})) = \hat{\omega}_1 \frac{\hat{\alpha}_n}{\alpha_n} \widehat{Z}_n^c(k_1) \rightsquigarrow \omega_1(Z_\infty(k_0) - Z_\infty(mk_0))^{-1} Z_\infty^c(k_1).$$

Similarly, $\hat{\omega}_2\hat{\alpha}_n(\hat{q}_n(\tau_{n2}) - q(\tau_{n2})) \rightsquigarrow \omega_2(Z_\infty(k_0) - Z_\infty(mk_0))^{-1} Z_\infty^c(k_2)$. Therefore,

$$\begin{aligned}
&\hat{\alpha}_n\Big[\hat{\omega}_1\hat{q}_n(\tau_{n1}) + \hat{\omega}_2\hat{q}_n(\tau_{n2}) - (\omega_1 q(\tau_{n1}) + \omega_2 q(\tau_{n2}))\Big] \\
=&\hat{\alpha}_n\Big[\hat{\omega}_1(\hat{q}_n(\tau_{n1}) - q(\tau_{n1})) + \hat{\omega}_2(\hat{q}_n(\tau_{n2}) - q(\tau_{n2}))\Big] + \hat{\alpha}_n\Big[(\hat{\omega}_1 - \omega_1)q(\tau_{n1}) + (\hat{\omega}_2 - \omega_2)q(\tau_{n2})\Big] \\
=&\frac{\omega_1 Z_\infty^c(k_1) + \omega_2 Z_\infty^c(k_2)}{Z_\infty(k_0) - Z_\infty(mk_0)} + o_p(1) + \hat{\alpha}_n\Big[(\hat{\omega}_1 - \omega_1)(q(\tau_{n1}) - q(1)) + (\hat{\omega}_2 - \omega_2)(q(\tau_{n2}) - q(1))\Big] \\
=&\frac{\omega_1 Z_\infty^c(k_1) + \omega_2 Z_\infty^c(k_2)}{Z_\infty(k_0) - Z_\infty(mk_0)} + o_p(1),
\end{aligned}$$

where the second and last equalities hold because $\hat{\omega}_1 + \hat{\omega}_2 = 1$ and $\hat{\alpha}_n(q(\tau_{nj}) - q(1)) = O_p(1)$ for $j = 1, 2$, respectively.

The last result holds because $\hat{\omega}_1 + \hat{\omega}_2 = 1$, $\alpha_n(q(1) - q(\tau_{n1})) = \eta(k_1)$, and $\alpha_n(q(1) - q(\tau_{n2})) = \eta(k_2)$.

C Proof of Theorem 4.1

We organize the proof into three steps. In the first step, we show

$$\hat{\alpha}_n\Bigg[\hat{\omega}_1\hat{q}_n(\tau_{n1})+\hat{\omega}_2\hat{q}_n(\tau_{n2})-q(1)\Bigg] \rightsquigarrow \frac{\omega_1 Z^c_\infty(k_1)+\omega_2 Z^c_\infty(k_2)}{Z_\infty(k_0)-Z_\infty(mk_0)}.$$

In the second step, we show that for $j=1$ and 2,

$$\hat{\alpha}_b(\hat{q}_b(\tau_{bj})-\hat{q}_n(\tau_{bj})) \rightsquigarrow Z^c_\infty(k_j).$$

Then, given the consistency of $\hat{\omega}_1$ and $\hat{\omega}_2$, we have

$$\tilde{Z}^*_b=\hat{\alpha}_b\Bigg[\hat{\omega}_1\bigg(\hat{q}_b(\tau_{b1})-\hat{q}_n(\tau_{b1})\bigg)+\hat{\omega}_2\bigg(\hat{q}_b(\tau_{b2})-\hat{q}_n(\tau_{b2})\bigg)\Bigg] \rightsquigarrow \frac{\omega_1 Z^c_\infty(k_1)+\omega_2 Z^c_\infty(k_2)}{Z_\infty(k_0)-Z_\infty(mk_0)},$$

which is the same as the limiting distribution of

$$\hat{\alpha}_n\Bigg[\hat{\omega}_1\hat{q}_n(\tau_{n1})+\hat{\omega}_2\hat{q}_n(\tau_{n2})-q(1)\Bigg].$$

In the third step, we show that the distribution of $\frac{\omega_1 Z^c_\infty(k_1)+\omega_2 Z^c_\infty(k_2)}{Z_\infty(k_0)-Z_\infty(mk_0)}$ is continuous. Therefore, the critical value $\hat{C}_{1-\alpha}$ of $\tilde{Z}^*_b$ is a consistent estimator of the critical value $C_{1-\alpha}$ of $\frac{\omega_1 Z^c_\infty(k_1)+\omega_2 Z^c_\infty(k_2)}{Z_\infty(k_0)-Z_\infty(mk_0)}$. This concludes the proof of Theorem 4.1.

Step 1:

Because $\hat{\xi} \xrightarrow{p} \xi$, ω_1 and ω_2 solve the following equations:

$$\omega_1+\omega_2=1 \quad \text{and} \quad \omega_1\eta(k_1)+\omega_2\eta(k_2)=0.$$

Therefore, by Corollary 3.1, we have

$$\hat{\alpha}_n\left[\hat{\omega}_1\hat{q}_n(\tau_{n1}) + \hat{\omega}_2\hat{q}_n(\tau_{n2}) - q(1)\right] \rightsquigarrow \frac{\omega_1 Z^c_\infty(k_1) + \omega_2 Z^c_\infty(k_2)}{Z_\infty(k_0) - Z_\infty(mk_0)}.$$

Step 2:

We note that

$$\hat{\alpha}_b(\hat{q}_b(\tau_{bj}) - \hat{q}_n(\tau_{bj})) = \hat{\alpha}_b(\hat{q}_b(\tau_{bj}) - q(\tau_{bj})) - \hat{\alpha}_b(\hat{q}_n(\tau_{bj}) - q(\tau_{bj})) = I + II.$$

We aim to show that

$$I \rightsquigarrow (Z_\infty(k_0) - Z_\infty(mk_0))^{-1} Z^c_\infty(k_j) \quad \text{and} \quad II \xrightarrow{p} 0.$$

For the first claim, let $P_{ni} = \sum_{l=1}^{b} \mathbb{1}\{I_{nl} = i\}$ where $(I_{n1}, \cdots, I_{nb})$ is a multinomial vector with parameter b and probability $(\frac{1}{n}, \cdots, \frac{1}{n})$ and

$$\widehat{Z}^c_b(k_j) = \alpha_b(\hat{q}_b(\tau_{bj}) - q(\tau_{bj}))$$

where $\alpha_b = 1/(q(1) - q(1 - 1/b))$. Then, following an argument similar to the proof of Theorem 3.1, we have

$$\widehat{Z}^c_b(k_j) = -\arg\min_z -\frac{1}{b}\sum_{i=1}^{n} P_{ni}\mathbb{1}\{X_i \le x\}kz + \sum_{i=1}^{n} P_{ni}\mathbb{L}(-\alpha_b(Y_i - q(1)), z)\mathbb{1}\{X_i \le x\}.$$

Lemma H.2 shows that, point-wise in z,

$$-\frac{1}{b}\sum_{i=1}^{n} P_{ni}\mathbb{1}\{X_i \le x\}kz \xrightarrow{p} -kp_0 z \tag{C.1}$$

and

$$\sum_{i=1}^{n} P_{ni}\mathbb{L}(-\alpha_b(Y_i - q(1)), z)\mathbb{1}\{X_i \leq x\} \rightsquigarrow \int \mathbb{L}(u, z)dN \tag{C.2}$$

in which $N(\cdot)$ is exactly the same Poisson random measure as defined in the proof of Theorem 3.1. Given (C.1) and (C.2), we can conclude that $\widehat{Z}_b^c(k_j) \rightsquigarrow Z_\infty^c(k_j)$. In addition, similar to the proof of Corollary 3.1, we can show that

$$\hat{\alpha}_b/\alpha_b \xrightarrow{p} (Z_\infty(k_0) - Z_\infty(mk_0))^{-1}.$$

Thus, we have established that

$$I \rightsquigarrow (Z_\infty(k_0) - Z_\infty(mk_0))^{-1} Z_\infty^c(k_j).$$

For term II, denote $f_x(y)$ as the conditional density of y given $X \leq x$. Then, by Theorem I.1, we have

$$(n/(1-\tau_{bj}))^{1/2} f_x(q(1-k_j/b))\Big(\hat{q}_n(\tau_{bj}) - q(\tau_{bj})\Big) = O_p(1).$$

Therefore,

$$\begin{aligned} II =& \hat{\alpha}_b\Big(\hat{q}_n(\tau_{bj}) - q(\tau_{bj})\Big) \\ =& \frac{\hat{\alpha}_b}{\alpha_b}\alpha_b\Big(\hat{q}_n(\tau_{bj}) - q(\tau_{bj})\Big) \\ =& O_p(1) \times O_p\Big(\frac{\alpha_b}{(nb)^{1/2} f_x(q(1-k_j/b))}\Big). \end{aligned}$$

Note that, $f_x^{-1}(q(\tau_n)) = q'(\tau_n)$. In addition, we claim, as $\tau \to 1$,

$$\frac{q'(\tau)(1-\tau)}{q(1) - q(\tau)} \to -\xi_0. \tag{C.3}$$

Note for $a > b$, $\frac{q(1-b(1-\tau)) - q(1-a(1-\tau))}{q(1)-q(\tau)} \to a^{-\xi_0} - b^{-\xi_0}$. By the monotonicity of the density,[6] we

[6] Here we implicitly assume the density is monotone decreasing. If it is monotone increasing, we can

have
$$\frac{q'(1-b(1-\tau))(a-b)(1-\tau)}{q(1)-q(\tau)} \geq \frac{q(1-b(1-\tau))-q(1-a(1-\tau))}{q(1)-q(\tau)}$$
$$\geq \frac{q'(1-a(1-\tau))(a-b)(1-\tau)}{q(1)-q(\tau)}.$$

Therefore,
$$\limsup_{\tau\to 1} \frac{q'(1-a(1-\tau))(1-\tau)}{q(1)-q(\tau)} \leq \frac{a^{-\xi_0}-b^{-\xi_0}}{a-b}.$$

Let $a=1$ and $b\uparrow 1$, we have
$$\limsup_{\tau\to 1} \frac{q'(\tau)(1-\tau)}{q(1)-q(\tau)} \leq -\xi_0.$$

Similarly, using another half of the inequality, we can show
$$\liminf_{\tau\to 1} \frac{q'(\tau)(1-\tau)}{q(1)-q(\tau)} \geq -\xi_0$$

and (C.3) holds. By (C.3),
$$\frac{\alpha_b}{(nb)^{1/2} f_x(q(1-k_j/b))} = \frac{q'(1-k_j/b)}{(q(1)-q(1-1/b))(nb)^{1/2}} \sim -\xi_0 k_j^{-1}(b/n)^{1/2} \to 0.$$

Step 3

We note that, for any $z\in\Re$,
$$P(Z^c_\infty(k)=z) \leq \sum_{h=1}^{\infty} P(-\mathcal{J}_h = z-\eta(k)) = 0.$$

So $Z^c_\infty(k)$ is continuous, and thus is $\frac{\omega_1 Z^c_\infty(k_1)+\omega_2 Z^c_\infty(k_2)}{Z_\infty(k_0)-Z_\infty(mk_0)}$.

simply reverse the order of the inequality. The results still hold.

D Proof of Theorem 4.2

Let $(z_{1n}, \cdots, z_{Ln})$ and $(z_1, \cdots, z_L)$ be in the joint supports of $(\tilde{Z}_n(k_1), \cdots, \tilde{Z}_n(k_L))$ and $(\tilde{Z}_\infty(k_1), \cdots, \tilde{Z}_\infty(k_L))$, respectively, such that

$$(z_{1n}, \cdots, z_{Ln}) \to (z_1, \cdots, z_L).$$

By Assumption 8, $f(z_1, \cdots, z_L; \xi, p)$ is continuous in all its arguments and $\pi(q(1) + v/\hat{\alpha}_n) \to \pi(q(1))$. Furthermore, $\rho(z-v)f(z_1 - v, \cdots, z_L - v; \xi, p)\pi(q(1)+v/\hat{\alpha}_n)$ is dominated by an integrable function of v. Therefore, by the dominated convergence theorm, we have, point-wise in z,

$$Q_n(z, z_{1n}, \cdots, z_{Ln}, \hat{\xi}, \hat{p}) \xrightarrow{p} Q_\infty(z, z_1, \cdots, z_L; \xi_0, p_0).$$

In addition, since $\rho(\cdot)$ is convex in z, so be $Q_n(\cdot)$ and $Q(\cdot)$. In view of Lemma H.1, we have verified (i) and assumed (ii) and (iii) in Assumption 8. Therefore, by Lemma H.1,

$$\theta_n^{BE}(z_{1n}, \cdots, z_{Ln}, \hat{\xi}, \hat{p}) \xrightarrow{p} \theta_\infty^{BE}(z, z_1, \cdots, z_L; \xi_0, p_0)$$

where $\theta_n^{BE}(\cdot)$ and $\theta_\infty^{BE}(\cdot)$ are defined in (4.4) and (4.6), respectively. Since the sequence $(z_{1n}, \cdots, z_{Ln})$ is arbitrary, we have

$$\theta_n^{BE}(z_1, \cdots, z_L, \hat{\xi}, \hat{p}) \xrightarrow{p} \theta_\infty^{BE}(z_1, \cdots, z_L, \xi_0, p_0)$$

uniformly over $(z_1, \cdots, z_L)$ in any compact subset of the joint support of $(\tilde{Z}_\infty(k_1), \cdots, \tilde{Z}_\infty(k_L))$.

In addition, we note that

$$\begin{pmatrix} \tilde{Z}_n(k_1) \\ \vdots \\ \tilde{Z}_n(k_L) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \tilde{Z}_\infty(k_1) \\ \vdots \\ \tilde{Z}_\infty(k_L) \end{pmatrix}.$$

Therefore, by the continuous mapping theorem,

$$\hat{Z}_n^{BE} \equiv \theta_n^{BE}(\tilde{Z}_n(k_1), \cdots, \tilde{Z}_n(k_L), \hat{\xi}, \hat{p}) \rightsquigarrow Z_\infty^{BE} \equiv \theta_\infty^{BE}(\tilde{Z}_\infty(k_1), \cdots, \tilde{Z}_\infty(k_L); \xi_0, p_0).$$

This concludes the proof.

E Proof of Theorem 4.3

The argument follows the proof of Chernozhukov and Hong (2004, Theorem 8). First, the proof of Theorem 4.2 implies

$$\theta_n^{BE}(z_1, \cdots, z_L, \hat{\xi}, \hat{p}) \xrightarrow{p} \theta_\infty^{BE}(z_1, \cdots, z_L; \xi_0, p_0),$$

where $\theta_\infty^{BE}(z_1, \cdots, z_L; \xi, p)$ is defined in (4.6). In addition,

$$f(z_1 - v, \cdots, z_L - v; \hat{\xi}, \hat{p}) \xrightarrow{p} f(z_1 - v, \cdots, z_L - v; \xi_0, p_0).$$

By Assumption 8,

$$\rho(\theta_n^{BE}(z_1, \cdots, z_L; \hat{\xi}, \hat{p}) - v) f(z_1 - v, \cdots, z_L - v; \hat{\xi}, \hat{p})$$

is dominated by an integrable function w.r.t. $(z_1, \cdots, z_L)$ for any $v \in K_t$ with t fixed, because $\rho(\cdot)$ increases at most polynomially in its argument, $\theta_n^{BE}(z_1, \cdots, z_L; \hat{\xi}, \hat{p})$ is polynomial in

$(z_1, \cdots, z_L)$, $f(\cdot; \xi, p)$ decreases exponentially in its arguments, and for every fixed t, K_t is compact. Therefore, by the dominated convergence theorem, as $n \to \infty$

$$\begin{aligned} &\int_{\Re^L} \rho(\theta_n^{BE}(z_1, \cdots, z_L; \hat{\xi}, \hat{p}) - v) f(z_1 - v, \cdots, z_L - v; \hat{\xi}, \hat{p}) dz_1 \cdots dz_L \\ \xrightarrow{p} &\int_{\Re^L} \rho(\theta_\infty^{BE}(z_1, \cdots, z_L; \xi_0, p_0) - v) f(z_1 - v, \cdots, z_L - v; \xi_0, p_0) dz_1 \cdots dz_L. \end{aligned} \tag{E.1}$$

By (4.6) and a change of variable argument, we have, for any v,

$$\theta_\infty^{BE}(z_1, \cdots, z_L; \xi_0, p_0) - v = \theta_\infty^{BE}(z_1 - v, \cdots, z_L - v; \xi_0, p_0).$$

Furthermore, by construction, $f(\cdot; \xi_0, p_0)$ is the joint PDF of $(\tilde{Z}_\infty(k_1), \cdots, \tilde{Z}_\infty(k_L))$. Therefore,

$$\begin{aligned} \text{the RHS of (E.1)} = &\int_{\Re^L} \rho(\theta_\infty^{BE}(z_1, \cdots, z_L; \xi_0, p_0)) f(z_1, \cdots, z_L; \xi_0, p_0) dz_1 \cdots dz_L \\ = &\mathbb{E}\rho(\theta_\infty^{BE}(\tilde{Z}_\infty(k_1), \cdots, \tilde{Z}_\infty(k_L); \xi_0, p_0)) = \mathbb{E}\rho(Z_\infty^{BE}), \end{aligned}$$

where the last equality holds because $Z_\infty^{BE} = \theta_\infty^{BE}(Z_\infty(k_1), \cdots, Z_\infty(k_L); \xi_0, p_0)$. Then, we have, for every fixed t,

$$\limsup_{n\to\infty} \int_{K_t} \int_{\Re^L} \rho(\theta_n^{BE}(z_1, \cdots, z_L; \hat{\xi}, \hat{p}) - v) f(z_1 - v, \cdots, z_L - v; \hat{\xi}, \hat{p}) dz_1 \cdots dz_L / \Lambda(K_t) = \mathbb{E}\rho(Z_\infty^{BE}).$$

Taking $\limsup_{t\to\infty}$ on both sides, we have

$$AAR_\rho(\{\theta_n^{BE}\}) = \mathbb{E}\rho(Z_\infty^{BE}).$$

To prove the second result, for each $t \geq 1$, we denote $\tilde{q}_{n,t}^{BE}$ as the Bayesian estimator with prior $\pi(\overline{q}) = \mathbb{1}\{\hat{\alpha}_n(\overline{q} - q(1)) \in K_t\}$, i.e.,

$$\hat{\alpha}_n(\tilde{q}_{n,t}^{BE} - q(1)) = \tilde{\theta}_t^{BE}(\tilde{Z}_n(k_1), \cdots, \tilde{Z}_n(k_1), \hat{\xi}, \hat{p})$$

where $\tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \xi, p)$ is defined in (4.7). Next, we show

$$\limsup_{t\to\infty} \limsup_{n\to\infty} AR_{\rho,K_t}(\tilde{\theta}_t^{BE}(\cdot; \hat{\xi}, \hat{p})) = \mathbb{E}\rho(Z_\infty^{BE}). \tag{E.2}$$

Following the proof of Theorem 4.2, for fixed t, as $n \to \infty$, we have

$$\tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \hat{\xi}, \hat{p}) \xrightarrow{p} \tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \xi_0, p_0)$$

uniformly over $(z_1, \cdots, z_L)$ in any compact subset of the joint support of $(\tilde{Z}_\infty(k_1), \cdots, \tilde{Z}_\infty(k_L))$. Therefore, as $n \to \infty$,

$$\begin{aligned}
&\rho(\tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \hat{\xi}, \hat{p}) - v) f(z_1 - v, \cdots, z_L - v; \hat{\xi}, \hat{p}) \\
\xrightarrow{p} &\rho(\tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \xi_0, p_0) - v) f(z_1 - v, \cdots, z_L - v; \xi_0, p_0).
\end{aligned}$$

In addition, by Assumption 8.5, for fixed t, $\rho(\tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \hat{\xi}, \hat{p}) - v) f(z_1 - v, \cdots, z_L - v; \hat{\xi}, \hat{p})$ is dominated by some function that is integrable w.r.t. $z_1, \cdots, z_L$ over $\Re^L$. Therefore, by the dominated convergence theorem,

$$\begin{aligned}
&\limsup_{n\to\infty} AR_{\rho,\lambda_t}(\tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \hat{\xi}, \hat{p})) \\
=& \int_{-t}^{t} \int_{\Re^L} \rho(\tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \xi_0, p_0) - v) f(z_1 - v, \cdots, z_L - v; \xi_0, p_0) dz_1 \cdots dz_L dv/2t \\
=& \int_{-1}^{1} \int_{\Re^L} \rho(\tilde{\theta}_t^{BE}(z_1, \cdots, z_L; \xi_0, p_0) - tu) f(z_1 - tu, \cdots, z_L - tu; \xi_0, p_0) dz_1 \cdots dz_L du/2 \\
=& \int_{-1}^{1} \int_{\Re^L} \rho(\tilde{\theta}_t^{BE}(z_1 + tu, \cdots, z_L + tu; \xi_0, p_0) - tu) f(z_1, \cdots, z_L; \xi_0, p_0) dz_1 \cdots dz_L du/2.
\end{aligned} \tag{E.3}$$

By the definition of $\tilde{\theta}_t^{BE}$,

$$
\begin{aligned}
&\tilde{\theta}_t^{BE}(w_1+tu,\cdots,w_L+tu;\xi_0,p_0)-tu\\
=&\arg\min_\gamma \int_{-t}^{t}\rho(\gamma+tu-v)f(w_1+tu-v,\cdots,w_L+tu-v;\xi_0,p_0)dv\\
=&\arg\min_\gamma \int_{\Re}\mathbb{1}\{v\in(t-tu,-t-tu)\}\rho(\gamma-v)f(w_1-v,\cdots,w_L-v;\xi_0,p_0)dv.
\end{aligned}
$$

Since $u\in(-1,1)$, as $t\to\infty$, $\mathbb{1}\{v\in(t-tu,-t-tu)\}\uparrow 1$. Therefore, by the bounded convergence theorem,

$$
\begin{aligned}
&\int_{\Re}\mathbb{1}\{v\in(t-tu,-t-tu)\}\rho(\gamma-v)f(w_1-v,\cdots,w_L-v;\xi_0,p_0)dv\\
\to&\int_{\Re}\rho(\gamma-v)f(w_1-v,\cdots,w_L-v;\xi_0,p_0)dv.
\end{aligned}
$$

Then, by Lemma H.1, as $t\to\infty$

$$
\tilde{\theta}_t^{BE}(z_1+tu,z_L+tu;\xi_0,p_0)-tu\to\theta_\infty^{BE}(z_1,\cdots,z_L;\xi_0,p_0). \tag{E.4}
$$

Following (E.3), in order to show (E.2), it suffices to show, as $t\to\infty$

$$
\begin{aligned}
&\int_{-1}^{1}\int_{\Re^L}\left|\rho(\tilde{\theta}_t^{BE}(z_1+tu,\cdots,z_L+tu;\xi_0,p_0)-tu)-\rho(\theta_\infty^{BE}(z_1,\cdots,z_L;\xi_0,p_0))\right|\\
&\times f(z_1,\cdots,z_L;\xi_0,p_0)dz_1\cdots dz_L du/2\\
=&\int_{-1}^{1}\int_{\Re^L}\left[\rho(\tilde{\theta}_t^{BE}(z_1+tu,\cdots,z_L+tu;\xi_0,p_0)-tu)-\rho(\theta_\infty^{BE}(z_1,\cdots,z_L;\xi_0,p_0))\right]^-\\
&\times f(z_1,\cdots,z_L;\xi_0,p_0)dz_1\cdots dz_L du/2\\
&+\int_{-1}^{1}\int_{\Re^L}\left[\rho(\tilde{\theta}_t^{BE}(z_1+tu,\cdots,z_L+tu;\xi_0,p_0)-tu)-\rho(\theta_\infty^{BE}(z_1,\cdots,z_L;\xi_0,p_0))\right]^+\\
&\times f(z_1,\cdots,z_L;\xi_0,p_0)dz_1\cdots dz_L du/2\\
=&I_t+II_t\to 0.
\end{aligned}
$$

For I_t, we have

$$\Big[\rho(\tilde{\theta}_t^{BE}(z_1+tu,\cdots,z_L+tu;\xi_0,p_0)-tu)-\rho(\theta_\infty^{BE}(z_1,\cdots,z_L;\xi_0,p_0))\Big]^- \leq \rho(\theta_\infty^{BE}(z_1,\cdots,z_L;\xi_0,p_0))$$

which, by Assumption 8.4, is integrable w.r.t. $f(z_1,\cdots,z_L;\xi_0,p_0)dz_1\cdots dz_L$. Therefore, by (E.4) and the dominated convergence theorem, we have $I_t \to 0$.

In addition, by (4.7),

$$\begin{aligned}
&\int_{-1}^{1}\int_{\Re^L}\rho(\tilde{\theta}_t^{BE}(z_1+tu,\cdots,z_L+tu;\xi_0,p_0)-tu)f(z_1,\cdots,z_L;\xi_0,p_0)dz_1\cdots dz_L du\\
=&\int_{-t}^{t}\int_{\Re^L}\rho(\tilde{\theta}_t^{BE}(z_1,\cdots,z_L;\xi_0,p_0)-v)f(z_1-v,\cdots,z_L-v;\xi_0,p_0)dz_1\cdots dz_L dv/2t\\
\leq&\int_{-t}^{t}\int_{\Re^L}\rho(\theta_\infty^{BE}(z_1,\cdots,z_L;\xi_0,p_0)-v)f(z_1-v,\cdots,z_L-v;\xi_0,p_0)dz_1\cdots dz_L dv/2t\\
=&\int_{-1}^{1}\int_{\Re^L}\rho(\theta_\infty^{BE}(z_1,\cdots,z_L;\xi_0,p_0))f(z_1,\cdots,z_L;\xi_0,p_0)dz_1\cdots dz_L du/2.
\end{aligned}$$

Therefore,

$$0 \leq II_t \leq I_t \to 0.$$

This concludes (E.2). If there exists a sequence of estimators, denoted as $\{\breve{\theta}_n\}$, such that $\breve{\theta}_n \in \Theta_n$ and it achieves strictly smaller asymptotic average risk than the Bayesian estimator θ_n^{BE}, then for infinitely many t and n,

$$AR_{\rho,K_t}(\breve{\theta}_n) < AR_{\rho,K_t}(\tilde{\theta}_t^{BE}(\cdot;\hat{\xi},\hat{p})).$$

This is a contradiction because, by construction,

$$\tilde{\theta}_t^{BE}(\cdot;\hat{\xi},\hat{p}) \in \arg\min_{\theta\in\Theta_n} AR_{\rho,K_t}(\theta).$$

This concludes the proof.

F Proof of Corollary 4.1

Denote $\hat{Z}_n^{BE}(\tau') = \hat{\alpha}_n(\hat{q}^{BE}(\tau') - q(1) = \theta_n^{BE}(\tilde{Z}_n(k_1), \cdots, \tilde{Z}_n(k_L); \hat{\xi}, \hat{p})$. Then we have

$$P(\hat{q}^{BE}(\tau') > q(1)) = P(\hat{Z}_n^{BE}(\tau') > 0) \to P(Z_\infty^{BE}(\tau') > 0).$$

Next, we show

$$P(Z_\infty^{BE}(\tau') > 0) = \tau'.$$

Suppose not, then there exists a nonzero constant c such that $P(Z_\infty^{BE}(\tau') > c) = \tau'$ or equivalently, by the first order condition,

$$\mathbb{E}\tilde{\rho}_{\tau'}(Z_\infty^{BE}(\tau') - c) < \mathbb{E}\tilde{\rho}_{\tau'}(Z_\infty^{BE}(\tau')).$$

Similar to the proof of the first result in Theorem 4.3, we can show $\mathbb{E}\tilde{\rho}_{\tau'}(Z_\infty^{BE}(\tau') - c)$ is the asymptotic average risk for the estimator $\theta_n^{BE}(\cdot; \hat{\xi}, \hat{p}) - c$, i.e.,

$$AAR_{\tilde{\rho}_{\tau'}}(\{\theta_n^{BE}(\cdot; \hat{\xi}, \hat{p}) - c\}) = \mathbb{E}\tilde{\rho}_{\tau'}(Z_\infty^{BE}(\tau') - c) < \mathbb{E}\tilde{\rho}_{\tau'}(Z_\infty^{BE}(\tau')) = AAR_{\tilde{\rho}_{\tau'}}(\{\theta_n^{BE}(\cdot; \hat{\xi}, \hat{p})\}).$$

On the other hand, $\theta_n^{BE}(\cdot; \hat{\xi}, \hat{p}) - c \in \Theta_n$. Therefore, we reach a contradiction to the second result in Theorem 4.3. This implies

$$P(Z_\infty^{BE}(\tau') > 0) = \tau'.$$

Then

$$P(\hat{q}^{BE}(\tau') \le q(1) \le \hat{q}^{BE}(\tau'')) \to P(Z_\infty^{BE}(\tau'') > 0) - P(Z_\infty^{BE}(\tau') > 0) = \tau'' - \tau'.$$

G Proof of Proposition 4.1

We consider the CDF evaluated at $(u_1, \cdots, u_L)$ such that $u_1 < u_2, \cdots, < u_L$. Note that

$$Z_\infty(k) = -\mathcal{J}_{h(k)} = -(\gamma_1^{h(k)}/p)^{-\xi},$$

where $\gamma_i^j = \sum_{l=i}^j \mathcal{E}_l$. Therefore,

$$\begin{aligned}
&P(\tilde{Z}_\infty(h(k_1)) \leq u_1, \cdots, \tilde{Z}_\infty(h(k_L)) \leq u_L) \\
=&\mathbb{E}P(\tilde{Z}_\infty(h(k_1)) \leq u_1, \cdots, \tilde{Z}_\infty(h(k_L)) \leq u_L | \gamma_1^{h(k_0)}, \gamma_1^{h(mk_0)}) \\
=&\mathbb{E}P\left(\frac{(\gamma_1^{h(k_1)})^{-\xi}}{(\gamma_1^{h(mk_0)})^{-\xi} - (\gamma_1^{h(k_0)})^{-\xi}} \leq u_1, \cdots, \frac{(\gamma_1^{h(k_L)})^{-\xi}}{(\gamma_1^{h(mk_0)})^{-\xi} - (\gamma_1^{h(k_0)})^{-\xi}} \leq u_L \middle| \gamma_1^{h(k_0)}, \gamma_1^{h(mk_0)} \right) \\
=&\mathbb{E}P\bigg(\gamma_{h(mk_0)+1}^{h(k_1)} \leq [u_1((\gamma_1^{h(mk_0)})^{-\xi} - (\gamma_1^{h(k_0)})^{-\xi})]^{-1/\xi} - \gamma_1^{h(mk_0)}, \cdots, \\
&\gamma_{h(mk_0)+1}^{h(k_L)} \leq [u_L((\gamma_1^{h(mk_0)})^{-\xi} - (\gamma_1^{h(k_0)})^{-\xi})]^{-1/\xi} - \gamma_1^{h(mk_0)} | \gamma_1^{h(k_0)}, \gamma_1^{h(mk_0)} \bigg)
\end{aligned} \tag{G.1}$$

Notice that

$$(\gamma_{h(mk_0)+1}^{h(k_1)}, \cdots, \gamma_{h(mk_0)+1}^{h(k_L)}) \perp\!\!\!\perp (\gamma_1^{h(k_0)}, \gamma_1^{h(mk_0)}).$$

Let $s = \gamma_1^{h(k_0)}$, $t = \gamma_{h(k_0)+1}^{h(mk_0)}$, $\tilde{u} = (t+s)^{-\xi} - s^{-\xi}$, respectively. Then,

$$\begin{aligned}
&\text{The RHS of (G.1)} \\
=&\int P\left(\gamma_{h(mk_0)+1}^{h(k_1)} \leq (u_1 \tilde{u}(t,s))^{-1/\xi} - t, \cdots, \gamma_{h(mk_0)+1}^{h(k_L)} \leq (u_L \tilde{u}(t,s))^{-1/\xi} - t \right) \\
&\times f_{h(k_0)}(s) f_{h(mk_0)-h(k_0)}(t) ds dt.
\end{aligned}$$

Take derivatives w.r.t. $(u_1, \cdots, u_L)$, we obtain that

$$
\begin{aligned}
& f(u_1, \cdots, u_L; \xi, p) \\
= & \int (-1/\xi)^L \tilde{u}(t,s)^{-L/\xi} \left[\prod_{l=1}^{L} u_l^{-1/\xi - 1} f_{h_l - h_{l-1}}(v_l - v_{l-1}) \right] f_{h(k_0)}(s) f_{h(mk_0) - h(k_0)}(t) ds dt,
\end{aligned}
$$

where $h_l = h(k_l)$ for $L \geq l \geq 1$, $h_0 = h(mk_0)$, $v_l = (u_l \tilde{u}(t,s))^{-1/\xi}$ for $L \geq l \geq 1$, and $v_0 = t$.

H Technical Lemmas

We first recall the convexity lemma attributed to Geyer (1996) and Knight (1999), which we use repeatedly in our proof.

Lemma H.1. *Suppose (i) a sequence of convex lower-semicontinuous functions Q_n: $\Re \mapsto \overline{\Re}$ marginally converges to Q_∞: $\Re \mapsto \overline{\Re}$ over a dense subset of $\Re$, (ii) Q_∞ is finite over a nonempty open set $\mathcal{Z}_0$, and (iii) Q_∞ is uniquely minimized at a random variable Z_∞, then any argmin of Q_n, denoted $\hat{Z}_n(1)$, converges in distribution of Z_∞.*

The following lemma is used to prove Theorem 4.1.

Lemma H.2. *If the assumptions in Theorem 4.1 hold, then*

$$
-\frac{1}{b} \sum_{i=1}^{n} P_{ni} \mathbb{1}\{X_i \leq x\} kz \xrightarrow{p} -kp_0 z
$$

and

$$
\sum_{i=1}^{n} P_{ni} \mathbb{L}(-\alpha_b(Y_i - q(1)), z) \mathbb{1}\{X_i \leq x\} \rightsquigarrow \int \mathbb{L}(u,z) dN.
$$

Proof. To establish the first result, we compute the characteristic function of $\frac{1}{b}\sum_{i=1}^{n} P_{ni}\mathbb{1}\{X_i \leq$

$x\}$. Let $\tilde{i}$ be the imaginary unit,

$$
\begin{aligned}
\mathbb{E}\exp(\frac{\tilde{i}t}{b}\sum_{i=1}^{n}P_{ni}\mathbb{1}\{X_i\le x\}) =&\Big[\mathbb{E}\exp(\frac{\tilde{i}t}{b}\sum_{i=1}^{n}\mathbb{1}\{I_{nl}=i\}\mathbb{1}\{X_i\le x\})\Big]^{b}\\
=&\Big[\mathbb{E}\exp(\frac{\tilde{i}t}{b}\mathbb{1}\{X_{I_{nl}}\le x\})\Big]^{b}\\
=&\Big\{1-\mathbb{E}\Big[\frac{\tilde{i}t}{b}\frac{b}{n}\sum_{l=1}^{n}\Big(1-\exp(\frac{1}{b}\mathbb{1}\{X_l\le x\})\Big)\Big]\Big\}^{b}
\end{aligned}
$$

Note that

$$
\frac{1}{n}\sum_{l=1}^{n}\Big(1-\exp(\frac{1}{b}\mathbb{1}\{X_l\le x\})\Big)=\Big(1-\exp(1/b)\Big)\Big(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_l\le x\}\Big)\xrightarrow{p}p_0.
$$

Therefore, by the dominated convergence theorem,

$$
\mathbb{E}\exp(\frac{\tilde{i}t}{b}\sum_{i=1}^{n}P_{ni}\mathbb{1}\{X_i\le x\})\to\exp(\tilde{i}tp_0)
$$

and thus $-\frac{1}{b}\sum_{i=1}^{n}P_{ni}\mathbb{1}\{X_i\le x\}kz\xrightarrow{p}-kzp_0$.

For the second part, denote $\hat{N}^*=\sum_{i=1}^{n}P_{ni}\mathbb{1}\{-\alpha_b(Y_i-q(1))\in\cdot,X_i\le x\}$. We have $\mathbb{E}P_{ni}=b/n$ and $\hat{N}\equiv\sum_{i=1}^{b}\mathbb{1}\{-\alpha_b(Y_i-q(1))\in\cdot,X_i\le x\}\rightsquigarrow N$ by the proof of Theorem 3.1. Then, by Resnick (2007, Proposition 6.2),

$$
P(\hat{N}^*\in\cdot|\text{Data})\xrightarrow{p}P(N\in\cdot).
$$

Taking expectation on both sides, we obtain that $\hat{N}^*\rightsquigarrow N$. Then, by the continuous mapping theorem, we have

$$
\int\mathbb{L}(u,z)d\hat{N}^*\rightsquigarrow\int\mathbb{L}(u,z)dN.
$$

□

I Estimations of the intermediate quantile and EV index

In this section, we consider the asymptotic properties of the estimator $\hat{q}_n(\tau_n)$ when $(1-\tau_n)n \to \infty$. In addition, we propose estimating the EV index by a Pickands-type estimator and show it is consistent. One example of intermediate τ_n, which is relevant to the subsampling inference proposed in Section 4, is $\tau_n = 1 - k_j/b$ for $j = 1$ and 2, because $(1-\tau_n)n = k_j n/b \to \infty$. We also need $\hat{\xi}$ when computing $\hat{\omega}_1$ and $\hat{\omega}_2$ in Theorem 4.1. The results in this section have already been established in the order statistics literature by Daouia et al. (2010). Since we approach the problem from the quantile regression perspective, we include these results for completeness. In addition, by clearly stating the assumptions for each result, we illustrate that our simulation-based inference methods do not require the second-order approximation, in contrast to the existing literature.

Assumption 10. *$\tau_n \to 1$ and $(1-\tau_n)n \to \infty$ polynomially in n.*

Assumption 11. *The conditional density of Y given $X \le x$ exists and is denoted as $f_x(\cdot)$. There exist some y_0 such that $f_x(\cdot)$ is monotonic for $y \ge y_0$.*

Theorem I.1. *Let $\lambda_n = n^{1/2} f_x(q(\tau_n))(1-\tau_n)^{-1/2}$. If Assumption 1, 4, 10, and 11 hold, then*

$$\lambda_n(\hat{q}_n(\tau_n) - q(\tau_n)) \rightsquigarrow \mathcal{N}(0, p_0^{-1}).$$

Proof. Let $\widehat{\Delta}_n = \lambda_n(\hat{q}_n(\tau_n) - q(\tau_n))$. Then we have

$$\widehat{\Delta}_n = \arg\min_{\Delta} -W_n\Delta + G_n(\Delta)$$

in which

$$W_n = (n(1-\tau_n))^{-1/2} \sum_{i=1}^{n} \mathbb{1}\{X_i \le x\}(\tau_n - \mathbb{1}\{Y_i \le q(\tau_n)\})$$

and

$$G_n = (n(1-\tau_n))^{-1/2} \sum_{i=1}^{n} \mathbb{1}\{X_i \le x\} \int_0^{\Delta} (\mathbb{1}\{Y_i \le q(\tau_n) + \frac{s}{\lambda_n}\} - \mathbb{1}\{Y_i \le q(\tau_n)\}) ds.$$

By the usual triangular array CLT, we have

$$W_n \rightsquigarrow \mathcal{N}(0, p_0).$$

Next, we aim to show that, point-wise in Δ, $G_n(\Delta) \xrightarrow{p} \frac{p_0\Delta^2}{2}$. Given this, by the convexity lemma in Pollard (1991), we have

$$\hat{\Delta}_n - p_0^{-1} W_n \xrightarrow{p} 0$$

and thus

$$\hat{\Delta}_n \rightsquigarrow \mathcal{N}(0, p_0^{-1}).$$

So it suffices to show $G_n(\Delta) \xrightarrow{p} \frac{p_0\Delta^2}{2}$. First, by the mean value theorem,

$$\begin{aligned}
\mathbb{E}G_n(\Delta) =& \frac{np_0}{\sqrt{n(1-\tau_n)}} \int_0^{\Delta} (F(q(\tau_n) + \frac{s}{\lambda_n}/x) - F(q(\tau_n)/x)) ds \\
=& \frac{p_0\Delta^2}{2} \frac{f_x(q(\tau_n) + \tilde{s}/\lambda_n)}{f_x(q(\tau_n))} \\
\to& \frac{p_0\Delta^2}{2}.
\end{aligned}$$

To see this, we first note that $\lambda_n = L_n (n(1-\tau_n))^{1/2}(1-\tau_n)^{\xi} \to +\infty$ where L_n is some slowly varying function and the EV index $\xi < 0$. In addition, because $n(1-\tau_n) \to \infty$, for any constant $l > 1$,

$$\frac{s}{\lambda_n} = \frac{s(q(l\tau_n) - q(\tau_n))}{(n(1-\tau_n))^{1/2}\int_1^l \frac{f_x(q(\tau_n))}{f_x(q(t\tau_n))}dt} \leq q(l\tau_n) - q(\tau_n).$$

If f_x is monotone increasing at its upper tail, then $\xi \geq -1$ and

$$\frac{f_x(q(\tau_n) + \tilde{s}/\lambda_n)}{f_x(q(\tau_n))} \in \left[\frac{f_x(q(\tau_n))}{f_x(q(\tau_n))}, \frac{f_x(q(l\tau_n))}{f_x(q(\tau_n))}\right] \to [1, l^{\xi+1}].$$

Since the above bounds hold for any $l > 1$, by letting $l \downarrow 1$, we have

$$\frac{f_x(q(\tau_n) + \tilde{s}/\lambda_n)}{f_x(q(\tau_n))} \to 1.$$

The case in which $f_x(\cdot)$ is monotone decreasing in its upper tail can be handled similarly.

Last, we show that $\mathrm{Var}(G_n(\Delta)) \to 0$.

$$\begin{aligned}
\mathrm{Var}(G_n(\Delta)) =& \frac{p_0}{1-\tau_n}\mathrm{Var}\left(\int_0^\Delta (\mathbb{1}\{Y_i \leq q(\tau_n) + \frac{s}{\lambda_n}\} - \mathbb{1}\{Y_i \leq q(\tau_n)\})ds\right) \\
\leq& \frac{p_0}{1-\tau_n}\mathbb{E}\left(\int_0^\Delta (\mathbb{1}\{Y_i \leq q(\tau_n) + \frac{s}{\lambda_n}\} - \mathbb{1}\{Y_i \leq q(\tau_n)\})ds\right)^2 \\
\leq& \frac{p_0\Delta^2}{(n(1-\tau_n))^{1/2}}\frac{f_x(q(\tau_n) + \tilde{\Delta}/\lambda_n)}{f_x(q(\tau_n))} \to 0.
\end{aligned}$$

This concludes that $G_n(\Delta) \xrightarrow{p} \frac{p_0\Delta^2}{2}$ and thus the proof. □

Next, we consider the Pickands-type estimator of the EV index as described in Resnick (2007, Section 4.5). For some positive integer R, $\{w_r\}_{r=1}^R$ is a set of weights which sum up to one. We estimate ξ by

$$\hat{\xi} = \sum_{r=1}^R \frac{-w_r}{\log(l)}\log\left(\frac{\hat{q}_n(ml^r\tau_n) - \hat{q}_n(l^r\tau_n)}{\hat{q}_n(ml^{r-1}\tau_n) - \hat{q}_n(l^{r-1}\tau_n)}\right),$$

in which τ_n is an intermediate order quantile index that satisfies Assumption 10, l and m are positive constants, and $\{\hat{q}_n(ml^r\tau_n), \hat{q}_n(l^r\tau_n)\}_{r=0}^R$ are computed based on (A.1) using the full sample.

Theorem I.2. *If the assumptions in Theorem I.1 hold, then*

$$\hat{\xi} \xrightarrow{p} \xi_0.$$

Proof. By Theorem I.1, we know that, for any intermediate order quantile index τ_n,

$$\hat{q}_n(\tau_n) - q(\tau_n) = O_p(\lambda_n^{-1}).$$

Therefore,

$$\begin{aligned}\frac{\hat{q}_n(ml^r\tau_n) - \hat{q}_n(l^r\tau_n)}{q_n(ml^r\tau_n) - q_n(l^r\tau_n)} =& 1 + \frac{\hat{q}_n(ml^r\tau_n) - q(ml^r\tau_n)}{q_n(ml^r\tau_n) - q_n(l^r\tau_n)} - \frac{\hat{q}_n(l^r\tau_n) - q(l^r\tau_n)}{q_n(ml^r\tau_n) - q_n(l^r\tau_n)} \\ =& 1 + O_p(((q_n(ml^r\tau_n) - q_n(l^r\tau_n))\lambda_n)^{-1}).\end{aligned}$$

In addition,

$$(q_n(ml^r\tau_n) - q_n(l^r\tau_n))\lambda_n = (n(1-\tau_n))^{1/2} \int_{l^r}^{ml^r} \frac{f_x(q(\tau_n))}{f_x(q(t\tau_n))} dt = O((n(1-\tau_n))^{1/2}).$$

So,

$$\begin{aligned}\frac{\hat{q}_n(ml^r\tau_n) - \hat{q}_n(l^r\tau_n)}{\hat{q}_n(ml^{r-1}\tau_n) - \hat{q}_n(l^{r-1}\tau_n)} \sim& (1 + O_p((n(1-\tau_n))^{-1/2}))\frac{q_n(ml^r\tau_n) - q_n(l^r\tau_n)}{q_n(ml^{r-1}\tau_n) - q_n(l^{r-1}\tau_n)} \\ \sim& (1 + O_p((n(1-\tau_n))^{-1/2}))l^{-\xi_0} \xrightarrow{p} l^{-\xi_0}.\end{aligned}$$

Then, by the continuous mapping theorem,

$$\hat{\xi} \xrightarrow{p} \sum_{r=1}^{R} w_r \xi_0 = \xi_0.$$

□

J Additional simulation results

Tables 4 and 5 report the coverage rates and average lengths of three procedures with $n = 10,000$.

Table 4: DGP 1, $n = 10,000$

	Sub		ABC			DFS	
	S1	S2	L=2 S1	L=2 S2	L=3 S1	S1	S2
$x = 1.1$	0.949	0.934	0.922	0.920	0.906	0.840	0.777
$np_0 \approx 1834$	(0.355)	(0.330)	(0.166)	(0.158)	(0.135)	(0.227)	(0.227)
$x = 2.2$	0.940	0.947	0.933	0.935	0.920	0.909	0.909
$np_0 \approx 3666$	(0.338)	(0.332)	(0.164)	(0.157)	(0.134)	(0.309)	(0.309)
$x = 3.3$	0.944	0.951	0.935	0.941	0.927	0.976	0.953
$np_0 \approx 5500$	(0.334)	(0.312)	(0.164)	(0.155)	(0.133)	(0.374)	(0.374)
$x = 4.4$	0.942	0.950	0.937	0.937	0.935	0.989	0.977
$np_0 \approx 7334$	(0.327)	(0.314)	(0.163)	(0.155)	(0.133)	(0.432)	(0.432)
$x = 5.5$	0.954	0.951	0.947	0.947	0.934	0.996	0.990
$np_0 \approx 9166$	(0.331)	(0.313)	(0.164)	(0.156)	(0.133)	(0.480)	(0.480)

Notes: The coverage rates and average lengths of the CIs (in parentheses) are reported.

Table 5: DGP 2, $n = 10,000$

	Sub		ABC			DFS	
	S1	S2	L=2 S1	L=2 S2	L=3 S1	S1	S2
$x = 3.3$	0.941	0.948	0.932	0.935	0.921	0.926	0.886
$np_0 \approx 3026$	(0.838)	(0.794)	(0.399)	(0.385)	(0.326)	(0.690)	(0.690)
$x = 4.4$	0.949	0.944	0.939	0.941	0.932	0.974	0.951
$np_0 \approx 5378$	(0.818)	(0.777)	(0.399)	(0.385)	(0.325)	(0.905)	(0.905)
$x = 5.5$	0.949	0.954	0.947	0.944	0.943	0.994	0.985
$np_0 \approx 8402$	(0.798)	(0.768)	(0.398)	(0.383)	(0.323)	(1.125)	(1.125)

Notes: The coverage rates and average lengths of the CIs (in parentheses) are reported.

Tables 6–9 report the finite sample performances of the point estimators from the three procedures. In particular, we consider the median-unbiased estimator for the subsampling approach, the posterior mean and median for the ABC approach, and the point estimator proposed by Daouia et al. (2010). Here we report the bias (BIAS), mean absolute deviation (MAD), and root mean squared error (RMSE).

Table 6: DGP 1, $n = 5,000$, Performances of point estimators

	Sub		ABC mean			ABC median			DFS	
	S1	S2	L=2 S1	L=2 S2	L=3 S1	L=2 S1	L=2 S2	L=3 S1	S1	S2
				$x = 2.2$			$np_0 \approx 1833$			
BIAS	0.005	0.002	0.015	0.013	0.011	0.004	0.002	0.003	0.016	0.018
MAD	0.067	0.062	0.050	0.046	0.044	0.047	0.043	0.042	0.091	0.103
RMSE	0.087	0.079	0.066	0.060	0.057	0.061	0.055	0.054	0.119	0.135
				$x = 3.3$			$np_0 \approx 2750$			
BIAS	0.002	0.000	0.013	0.010	0.007	0.002	0.000	0.000	0.011	0.012
MAD	0.064	0.006	0.047	0.044	0.038	0.044	0.042	0.037	0.088	0.102
RMSE	0.082	0.078	0.061	0.057	0.049	0.057	0.053	0.046	0.115	0.133
				$x = 4.4$			$np_0 \approx 3667$			
BIAS	0.001	0.000	0.012	0.010	0.008	0.001	0.000	0.000	0.008	0.009
MAD	0.064	0.059	0.047	0.043	0.040	0.044	0.040	0.039	0.089	0.101
RMSE	0.081	0.074	0.060	0.055	0.051	0.055	0.051	0.049	0.114	0.130
				$x = 5.5$			$np_0 \approx 4583$			
BIAS	0.000	0.000	0.013	0.010	0.007	0.002	0.000	0.000	0.009	0.011
MAD	0.062	0.059	0.045	0.043	0.038	0.042	0.037	0.037	0.088	0.102
RMSE	0.079	0.075	0.059	0.055	0.049	0.054	0.047	0.047	0.144	0.131

Table 7: DGP 2, $n = 5,000$, Performances of point estimators

	Sub		ABC mean			ABC median			DFS	
	S1	S2	L=2 S1	L=2 S2	L=3 S1	L=2 S1	L=2 S2	L=3 S1	S1	S2
				$x = 3.3$			$np_0 \approx 1513$			
BIAS	0.008	0.007	0.035	0.036	0.025	0.008	0.006	0.006	0.047	0.054
MAD	0.170	0.155	0.127	0.118	0.112	0.120	0.110	0.108	0.227	0.259
RMSE	0.221	0.202	0.166	0.154	0.146	0.154	0.141	0.139	0.305	0.347
				$x = 4.4$			$np_0 \approx 2689$			
BIAS	0.000	0.000	0.030	0.028	0.018	0.003	0.000	0.000	0.026	0.031
MAD	0.158	0.147	0.116	0.109	0.101	0.110	0.102	0.098	0.217	0.249
RMSE	0.201	0.189	0.150	0.140	0.128	0.139	0.140	0.123	0.280	0.324
				$x = 5.5$			$np_0 \approx 4201$			
BIAS	0.007	0.002	0.033	0.029	0.021	0.006	0.001	0.002	0.022	0.028
MAD	0.155	0.146	0.114	0.108	0.097	0.107	0.101	0.094	0.217	0.248
RMSE	0.201	0.188	0.145	0.139	0.124	0.135	0.128	0.118	0.278	0.317

Table 8: DGP 1, $n = 10,000$, Performances of point estimators

	Sub		ABC mean			ABC median			DFS	
	S1	S2	L=2 S1	L=2 S2	L=3 S1	L=2 S1	L=2 S2	L=3 S1	S1	S2
				$x = 1.1$			$np_0 \approx 1834$			
BIAS	0.002	0.002	0.011	0.009	0.007	0.003	0.001	0.002	0.012	0.014
MAD	0.046	0.045	0.035	0.033	0.031	0.032	0.031	0.030	0.065	0.075
RMSE	0.061	0.058	0.046	0.043	0.041	0.042	0.040	0.039	0.088	0.101
				$x = 2.2$			$np_0 \approx 3666$			
BIAS	0.001	0.000	0.009	0.008	0.006	0.001	0.000	0.000	0.007	0.009
MAD	0.046	0.043	0.033	0.031	0.029	0.031	0.029	0.028	0.064	0.073
RMSE	0.058	0.055	0.043	0.040	0.037	0.040	0.037	0.035	0.082	0.094
				$x = 3.3$			$np_0 \approx 5500$			
BIAS	0.000	0.000	0.009	0.007	0.006	0.000	0.000	0.000	0.005	0.006
MAD	0.045	0.042	0.032	0.030	0.028	0.030	0.028	0.027	0.063	0.072
RMSE	0.057	0.054	0.042	0.038	0.035	0.038	0.036	0.034	0.081	0.093
				$x = 4.4$			$np_0 \approx 7334$			
BIAS	0.009	-0.002	0.008	0.006	0.005	0.000	-0.002	0.000	0.007	0.008
MAD	0.045	0.041	0.032	0.030	0.027	0.030	0.029	0.026	0.063	0.072
RMSE	0.057	0.053	0.041	0.039	0.035	0.038	0.036	0.033	0.081	0.093
				$x = 5.5$			$np_0 \approx 9166$			
BIAS	0.000	0.000	0.009	0.008	0.041	0.001	0.000	0.000	0.006	0.006
MAD	0.044	0.041	0.032	0.030	0.027	0.030	0.028	0.026	0.063	0.072
RMSE	0.055	0.052	0.041	0.038	0.034	0.038	0.035	0.033	0.080	0.091

Table 9: DGP 2, $n = 10,000$, Performances of point estimators

	Sub		ABC mean			ABC median			DFS	
	S1	S2	L=2 S1	L=2 S2	L=3 S1	L=2 S1	L=2 S2	L=3 S1	S1	S2
				$x = 3.3$			$np_0 \approx 3026$			
BIAS	-0.004	-0.001	0.018	0.019	0.011	0.000	-0.001	-0.002	0.017	0.021
MAD	0.110	0.104	0.081	0.076	0.070	0.076	0.072	0.068	0.153	0.177
RMSE	0.140	0.133	0.104	0.098	0.090	0.097	0.090	0.086	0.197	0.228
				$x = 4.4$			$np_0 \approx 5378$			
BIAS	-0.002	0.000	0.019	0.019	0.018	0.000	0.000	-0.002	0.012	0.014
MAD	0.108	0.104	0.079	0.075	0.067	0.075	0.070	0.065	0.154	0.177
RMSE	0.138	0.132	0.101	0.096	0.085	0.094	0.088	0.082	0.198	0.229
				$x = 5.5$			$np_0 \approx 8402$			
BIAS	0.000	0.000	0.018	0.018	0.009	-0.001	-0.002	-0.004	0.012	0.012
MAD	0.108	0.100	0.076	0.073	0.065	0.072	0.067	0.063	0.156	0.179
RMSE	0.137	0.127	0.098	0.093	0.082	0.091	0.086	0.078	0.197	0.226